CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE
Faculty of Tropical AgriSciences

Genetic distances among dog breeds

MASTER'S THESIS

Prague 2019

Author: Bc. Petr Matějů

Chief supervisor: Mgr. Barbora Černá Bolfíková, Ph.D.

Declaration

I hereby declare that I have done this thesis entitled Genetic distances among dog breeds independently, all texts in this thesis are original, and all the sources have been quoted and acknowledged by means of complete references and according to Citation rules of the FTA.

In Prague 26.4.2019

....

Petr Matějů

Acknowledgements

Abstract

Dogs are the first domesticated animals and they accompanied people from the era of hunters and gatherers. However, modern dog breeds are usually not older than 200 years. Genomic studies revealed relationships and history of many of those breeds. Aim of this study was to differentiate each breed according to the genotypic data and to assign each breed into its private cluster. We used combination of 27 nuclear microsatellite markers. We separated the dog breeds into nine groups according to latest genomic studies. Our markers and analytical methods provided comparable clustering resolution as known from the published studies. Moreover, we were able to observe genetic differentiation that was not discovered before. New structures were found in varieties of German Pointers, Dutch Shepherds, Collies and Weimaraners. Our findings uncover structure in Dachshunds with origin in the Czech Republic and Slovakia. Our data indicate that the origin of Polish Greyhound is not as ancient as it is believed. Between breeds with high gene flow no structure was found. Two clustering programs Snapclust and STRUCTURE were compared. Snapclust performed faster and better in smaller datasets. STRUCTURE outperformed Snapclust in larger datasets of more closely related breeds.

Key words: Dog, breed assignment, microsatellite, genetic distance

Contents

List of tables

List of figures

1.   Introduction and Literature Review

1.1.   Domestication

Domestic dog (*Canis lupus familiaris*) is the sole member of the *Canidae* family only large carnivore and the first animal ever to be domesticated (Clutten-Brock 1995; Freedman & Wayne 2017). In the 20th century there were discussions whether the only ancestor of domestic dog is gray wolf (*Canis lupus*) or if golden jackal (*Canis aureus*) and coyote (*Canis latrans*) did not help form the new emerging species since all can hybridize (Clutten-Brock 1995). Today it is clear and proven by many studies including archaeology, genetics, behavioral studies that the only contributor to the dog's gene pool is the gray wolf (Clutten-Brock 1995; Lindblad-Toh et al. 2005).

Time of divergence between dogs and wolves is well discussed. Assumptions made by Skoglund et al. (2015) estimated the earliest split between dogs and wolves to early upper Paleolithic to 27,000 - 40,000 years before present (BP). The analysis was based on 16S rRNA isolated from Taimyr 1 male wolf individual excavated at Taimyr lake in north Siberia, with the support of radiocarbon dating. Statistical analyses date the Taimyr wolf very close to the split between the dog and wolf lineages. Wang et al. (2016) dated dog origin to 33,000 years BP using genome wide SNPs (Single Nucleotide Polymorphisms). Mathematical model by Lindblad-Toh et al. (2005) dated divergence approximately 9,000 generations ago, which would mean 27,000 years BP. Other studies focused on archaeology found "dog-like canids" matched this time period (Germonpré et al. 2009; Ovodov et al. 2011). However archaeological excavations confidently assigned to dogs are not older than 15,000 years BP in Europe and 12,500 years BP in Far East (Larson et al. 2012). Because domestication is a complex and prolonged process, it is not easy to estimate its exact timing. (Larson et al. 2012; Freedman et al. 2014; Wang et al. 2016). Molecular data obtained by Freedman et al. (2014) estimated the divergence of dogs from wolves between 11,000 years BP to 16,000 years BP. Even the earliest estimates date dog domestication in the era before humans transitioned from hunters-gatherers to agriculture and farming, making dog the only animal domesticated before humans settled down (Larson et al. 2012; Freedman et al. 2016). In history, dogs followed humans to all continents except Antarctica (Clutten-Brock 1995).

Many studies also address the original area of domestication and number of founding events. Until recently, many authors agreed on single origin but disagreed on the location (Freedman et al. 2014; Frantz et al. 2016). Based on maternally transmitted mitochondrial DNA and its control region sequence Pang et al. (2009) in their work suggested origin of dogs in area South of Yangtze river in China. Similar claim was later supported by Wang et al. (2016). Early SNP genotyping by vonHoldt et al. (2010) found high haplotype sharing between Middle Eastern wolf population and dogs suggesting the primary source of genetic variation there. Thalmann et al. (2013) recovered ancient mitochondrial DNA samples and after comparing them with samples of contemporary wolves and dogs suggested the origin of domestication in Europe. On the other hand, Shannon et al. (2015) after analysing of the Y chromosome put the place of origin to central Asia. Latest study of Frantz et al. (2016) is suggesting dual independent origin of dogs. Research with the highest resolution so far is combining genomics, archaeological data and ancient mtDNA sequences. This research stated a hypothesis of two independent domestication events - one in Western Eurasia and one in Central Asia (Figure 1). Frantz et al. (2016) also hypothesised, that the Asian dog population followed human dispersal to the west in between 6,400 – 14,000 years BP. After reaching Europe these dogs partially replaced native dog population (Figure 2). Human movement via migrations or trading increased dog gene flow and homogenized the populations. That lead to interbreeding between lineages, and therefore blurring their genetic signature. This fact complicates attempts to find the exact location of the dog's origin (Larson et al. 2012).
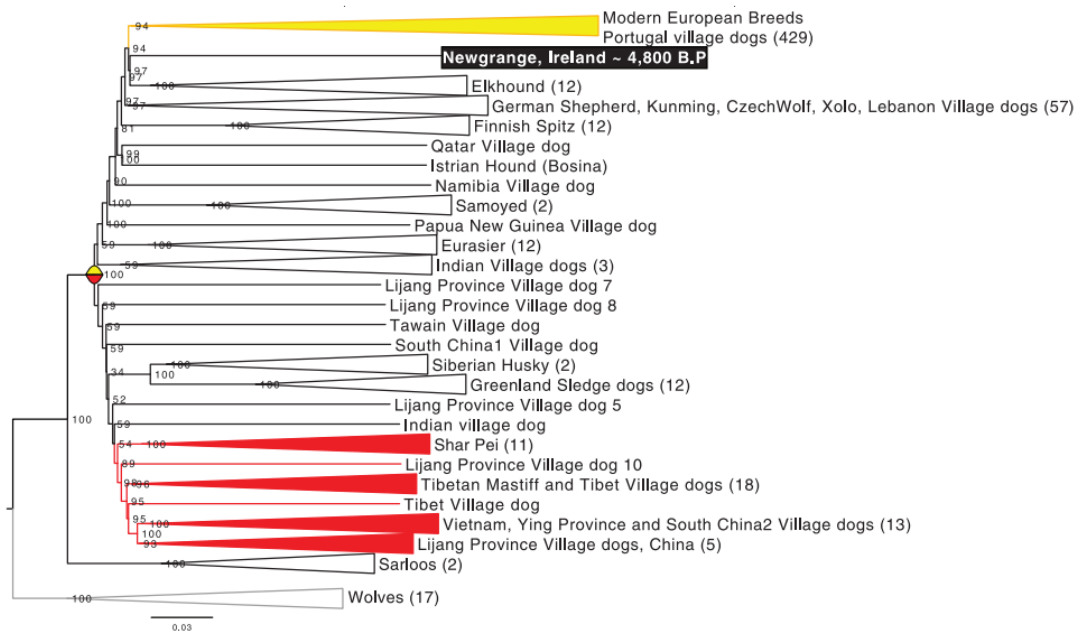
Figure 1: A neighbor-joining tree (with bootstrap values) based on identity by state (12) of 605 dogs from Frantz et al. (2016). Apparent split between Red and yellow clades representing the East Asian and Western Eurasian core groups.
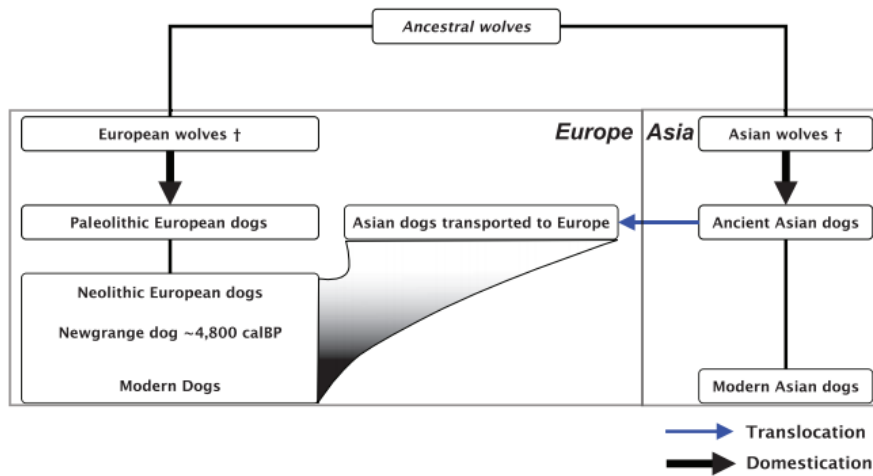
Figure 2: Dual origin model of dog domestication by Frantz et al. (2016) describes two independent domestication events from wolf populations which are now both extinct. Blue arrow is showing migration of Asian dogs westwards causing partial replacement of West Eurasian dog population.

The dog-wolf divergence is often described as a two-staged process. In the first phase humans and wolves occupied the same niche. Eventually both species evolved mutualistic behaviour, where wolves or so called proto-dogs and men cooperated in hunt or defence against other humans or carnivores (Thalmann et al. 2013; Freedman & Wayne 2017). There are even theories that this cooperation could have contributed to the

3

Neanderthals' extinction (Shipman 2015). This mutualistic behaviour enhanced when the first long-distance weapons like arrows were first used in the Mesolithic period. Dogs helped to track the shot-wounded animal and retrieved it back to the humans. This kind of teamwork thus increased the hunting efficiency (Clutten-Brock 1995; Freedman & Wayne 2017). It is believed that these first dogs did not descend from modern wolves, but rather already extinct Late Pleistocene wolf populations (Thalmann et al. 2013; Freedman et al. 2014; Freedman & Wayne 2017). In the second phase, around 10,000 years BP, humans started with farming and agriculture. This change of lifestyle enabled dogs to be used for other purposes like herding or companionship and not only for the hunt. The dogs could be selected for wider traits than during the hunters-gatherers period. This was the start of the vast phenotypic differentiation among dogs (Lindblad-Toh et al. 2005; Wang et al. 2016; Freedman & Wayne 2017).

1.2.   Early breed development

The first records of primitive dog breeds are on the Egyptian artworks dated to 6,500 years BP. Drawings depict a long legged-thin bodied "sighthound" type of dog. On the other hand, a small shaggy curly tailed dog is pictured on old Sumerian jewellery. Later, around 3,000 years BP, there is an evidence of "mastiff" type dogs in the fertile crescent, "spitz" type dogs in Far East or light fast types of dogs around the Mediterranean (Parker 2012). Excavations at Vindolanda fort complex located 3 km south of Hadrian's Wall in northern England inhabited between 1,500 years BP to 1,000 years BP yielded 520 bones of the domestic dogs showing high variance in size and skull morphology (Figure 3) (Bennett et al. 2016). Later, in the year 1087, on British Islands the Forest Laws of William the conqueror restricted commoners to own hounds or spaniels allowing them only to have a guarding mastiff type of dogs and "little dogs" as companions to prevent poaching in the Landlord's forests. This rule enforced reproductive isolation between the breeds to avoid punishment for owning a restricted dog type. These laws may have started human desire for phenotype selection, since commoners possibly tried to crossbreed dogs for the hunt, so they do not visually resemble any of the specifically forbidden breeds (Parker 2012).
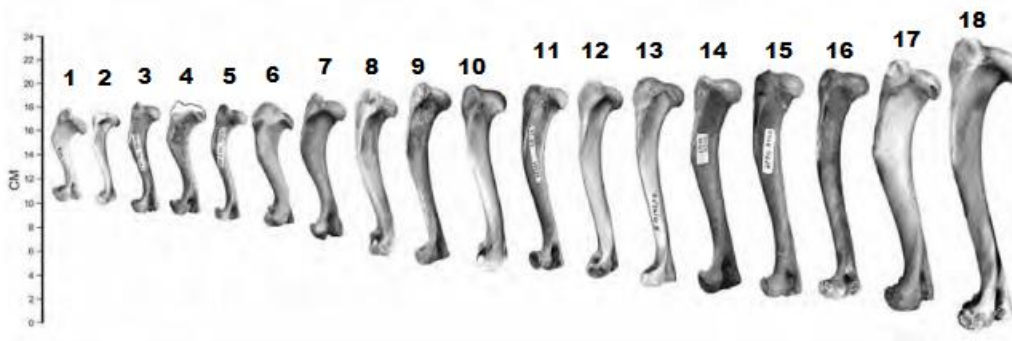
Figure 3: Dog humeri from Vindolanda in size comparison to modern dogs and a wolf. 1 – Dachshund, 2 – German Spitz (Pomeranian), 3 – Vindolanda, 4 – Vindolanda, 5 – Vindolanda, 6 – Scottish Terrier, 7 – Vindolanda, 8 – English Cocker Spaniel, 9 – Vindolanda, 10 – Chow-Chow, 11 – Vindolanda, 12 – Dingo, 13 – Greyhound, 14 – Vindolanda, 15 – Vindolanda, 16 – Vindolanda, 17 – Great Pyrenees, 18 – Wolf (Bennett et al. 2016)

Some of the ancient breeds, mentioned above, did remain to this date. VonHoldt et al. (2010) after analyzing 48,000 SNPs divided ancient breeds to Asian group (Dingo, New Guinea singing dog, Chow-chow, Akita, Chinese Shar-pei), Middle Eastern group containing two sighthounds (Afghan, Saluki), Northern group with Alaskan malamute and Siberian husky) and Basenji as African dog. These findings are supported with minor differences by other authors (Parker et al. 2004; Larson et al. 2012). Probably, one of the most ancient breed is African Basenji. Analysis done by Freedman et al. (2014) suggests Basenji − Middle Eastern wolf gene flow. Whether it was admixture from wolves to Basenjis or the other way around is not clear. However, it did enhance overall genetic similarity between these two populations. To the contrary some of the commonly believed most ancient breeds like the Rhodesian ridgeback or the Pharaoh hound were proven to separate from other breeds later than was thought. This means a possible re-creation of the breed by recent admixture of already existing breeds in order to match their ancestral appearance (Parker et al. 2004; Boyko et al. 2009).

Until 19th century dogs' gene pool was rather homogenous. This homogeneity was caused by history of human migrations. Most of all 400 plus breeds were created just recently, in the past two centuries (Leroy et al. 2009; Larson et al. 2012). During the Victorian era, dogs' selection for a specific appearance increased on popularity. In that time, the desire for fancy and novelty traits resulted in creation of a broad range of various dog phenotypes (Freedman & Wayne 2017). During this time period, first breed clubs and kennel clubs were founded. The first state level organization was "The Kennel Club"

in the United Kingdom (1873), followed by American Kennel Club (AKC) in 1884. The first breed club was the "Bulldog Club Incorporated" established in 1875 (DDTC 2018).

In order to improve the dogs' phenotype, breed standards were implemented to regulate shape, size colour or sometimes even behaviour. Animal who did not meet the internal requirements could had been restricted to breed. Keeping closed studbooks enforced the breed barrier rule. This rule states that "no dog may become a registered member of a breed unless its dam and sire are registered members". This rule basically ensures that each breed is forming its own isolated population (Parker 2012). Sometimes only few individuals contributed to the whole genepool of the newly forming breed, either because of their unique traits, wolf-dog mixing or breed rescue (e.g. Saarloos wolfdog, Irish wolfhound or Biewer Terrier) (BBCA 2014; Saarloos Wolfdog Club UK & Érie 2017; Urfer 2009). These processes imply that the breeds' gene pools were influenced by founder effect that comes along with bottlenecks. New challenges arise as on average, less than 10 % of registered dogs in a breed contributes to the next generations. Even that the population size may seem high, the effective population size is severely reduced. The trend of using popular sires and availability of frozen semen deepens this issue even further. When the current generation lacks specific traits valued by the breeders, the frozen semen even from already deceased animals which had the phenotype can be used. This individual can therefore contribute to the breed's gene pool time and time again (Parker 2012). This selective pressure on sires, bottlenecks and founder effects are causing majority of the breeds' genetic diversity loss (Lindblad-Toh et al. 2005; vonHoldt et al. 2010; Marsden et al. 2016). Over the centuries, these processes created observable genetical distances between breeds (Irion et al. 2003; Parker et al. 2004; Leroy et al. 2009; vonHoldt et al. 2010; Mellanby et al. 2013; Parker et al. 2017).

1.3.   Genetic distances observed

In the past two decades, scientists tried several approaches to compare between-breed relationships of domestic dogs. Different sets of genetic markers or statistical methods were used with various successes in breed clustering or phylogeny (Irion et al. 2003; Parker et al. 2004; Nicholas et al. 2008; Leroy et al. 2009; vonHoldt et al. 2010; Parker et al. 2017).

In the first decade of the 21st century most of the studies were done on autosomal microsatellites (Irion et al. 2003; Koskinen 2003; Parker et al. 2004; Schelling et al. 2005; Veit-Kensch et al. 2007; Leroy et al. 2009). These markers are easy to use, widely available, have high polymorphism and do not discriminate between sexes (Irion et al. 2003). It is however important to choose the right microsatellites for the research. Some markers can have too high mutation rate, others being monomorphic or in linkage disequilibrium. These errors mentioned above, make certain microsatellites not suitable for population studies and these markers must be excluded from the research (examples seen in (Table 1) (Irion et al. 2003; Schelling et al. 2005; Veit-Kensch et al. 2007).

Author	Nmic total2	Nmic Excluded	Reason
Irion et al. (2003)	100	34	Extremely Polymorphic
Schelling et al. (2005)	25	4	Monomorphic; linkage-disequilibrium
Veit-Kensch et al. (2007)	23	3	Extremely Polymorphic, unstable PCR results

Table 1: Microsatellite loci exclusion from published studies.

Other authors used genomic SNPs to study dog breeds' relationships (Lindblad-Toh et al. 2005; vonHoldt et al. 2010; Larson et al. 2012; Frantz et al. 2016; Parker et al. 2017). SNPs provide higher resolution since the number of markers can reach up to hundreds of thousands. However, requirement of expensive laboratory equipment can be the limiting factor of these markers.

1.3.1.   Microsatellites

In search of breeds' relatedness and relationships, research teams mainly used Bayesian analytical approaches. This method is often used to assign individuals to clusters according to their genetical fingerprint. Studies varies in number of loci used, number of breeds included in the research and sampled number of individuals per breed.

Over all, it is apparent that in all studies the success rate of assigning individuals to their own breed while using microsatellite markers exceeded 86 % (Table 2) with an exception of Mellanby et al. (2013) who had only Cavalier King Charles Spaniels pedigreed and other dogs were assigned to a breed by owner report or phenotype based by veterinarian observation. With the usage of SNPs, the assignment rate was 100%. It is therefore apparent that the main genetic structure of domestic dog population worldwide is based on the breed distances. Only exception is when the breed is very new and yet does not differ much from its parental populations. Parson Jack Russel and Jack Russel Terrier can be used as an example. Mellanby et al. (2013) points out that some Jack Russel Terriers bred in the UK meet the morphological criteria to by registered as Parson Jack Russel under AKC. Other studies that included Russel Terriers support this (Leroy et al. 2009; Parker et al. 2017).

Author	N_b	N_{ind}	N_{mic}	Success Rate
Koskinen 2003	5	50	10	100%
Irion et al. 2003	28	29 – 45	10	N/A
Parker et al. 2004	85	4 – 5	96	99%
Schelling et al. 2005	7	16 – 69	26	96,5%
Parker et al. 2007	132	4 – 5	96	N/A
Veit-Kensch et al. 2007	9*	12 – 33	23	99%
Leroy et al. 2009	61	20 – 30	21	86 - 98 %

Table 2: Overview of material used in microsatellite studies and their assignment success rate. N_b - Number of breeds, N_{ind} - Number of individuals, N_{mic} - Number of microsatellite loci. * - Research contained 9 dog breeds and a Polish wolf

Throughout the studies, visible patterns appeared. Some breeds or breed groups tended to cluster together. Parker et al. (2004) did a broad research on American dog population. Their data from 96 microsatellite markers were processed in STRUCTURE program utilizing Bayesian inference based on Markov Chain Monte Carlo repetitions (Pritchard et al. 2000). Clusters that divided at clusters (K) = 2 - 4 outlined breed grouping. First cluster that split at K = 2 was Ancient-Asian group, containing very old breeds like Basenji, Akita or Chow Chow. Second split was observed at K = 3. Mastiff like dogs including Bulldog, Boxer, and their close relatives separated from the rest. Another separation was visible at K = 4 where herding dogs and sighthounds formed their own distinct bracket. The rest of the dogs were classified by Parker as the "hunting group". Parker et al. (2007) provided additional data where they uncovered one more

cluster. So called "mountain group" splits at K = 5 from Hunting group. It is anchored with Bernese mountain dog and Greater Swiss mountain dog. In K > 5 more structures appeared but were inconsistent throughout the runs.

Study of Leroy et al. (2009) observed similar trends. Samples of 1514 dogs out of 61 dog breeds (25 dogs per breed on average) collected in France were tested on 21 microsatellite markers recommended by the International Society for Animal Genetics (ISAG). Structure analysis at K = 2-7 showed similar grouping patterns to ones described by (Parker et al. 2004, 2007). With K = 61 program correctly assigned 50 breeds to their individual clusters (Figure 4). Cursinu did not form a private cluster and even kept failing in assignment with just 18 % while using Structure and 86 % while using the method of direct assignment. This breed had to be therefore removed from further analysis. Some breeds clustered together, for example Czechoslovakian wolfdog with German Shepherd or Cavalier King Charles Spaniel with King Charles Spaniel. This is because STRUCTURE software detects only high hierarchical organization levels and as the number of breeds analysed in the studies is large, the program did not manage to separate breeds that were relatively similar (Leroy et al. 2009). The pairs that clustered together are generally new breeds. The Kennel Clubs decide if a breed variety will get the status of official breed. These new breeds can be therefore either newly registered varieties or a newly recognised crossbreed of two different breeds. These forms are genetically close to each other, that STRUCTURE does not recognise any pattern and clusters them together. This pattern was also present in other studies. Parker et al. (2007) achieved individual clustering in 112 breeds out of 132. Examples of breed pairs clustering together in this study are Alaskan Malamute with Siberian Husky, Petit Basset Griffon Vendeen with Grand Basset Griffon Vendeen, Belgian Sheepdog and Belgian Tarvuren. For the pairs mentioned above a separated run was necessary to achieve individual clustering. Some CSV individuals were assigned incorrectly as Saarloos Wolfdog in Leroy et al. (2009), therefore Czechoslovakian wolfdog – German Shepherd - Saarloos wolfdog triplet was analyzed separately. In all cases, except Belgian Sheepdog and Belgian Tervuren in Parker et al. (2004) and Petit Basset Griffon Vendeen and Grand Basset Griffon Vendeen in Parker et al. (2007), the analysis achieved correct separation between breeds. Parson Jack Russel and Jack Russel terrier were included as "one breed" in Leroy et al. (2009).
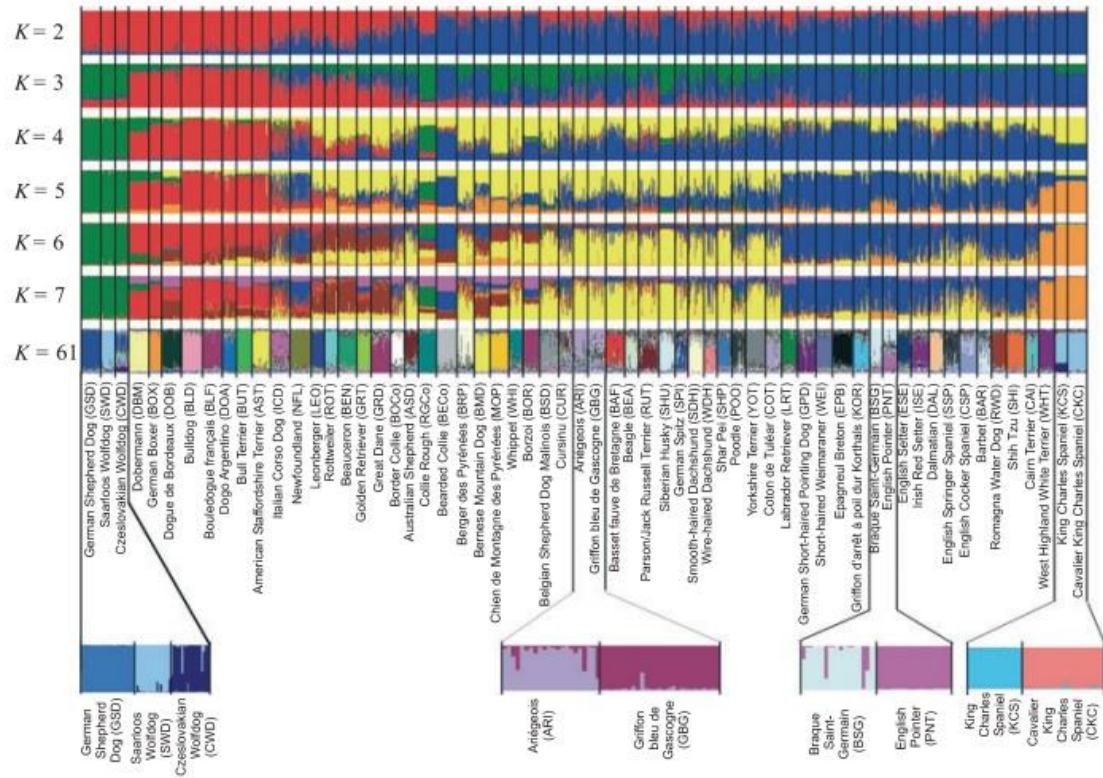
Figure 4: Structure analysis of 61 dog breeds. Runs K = 2 - 7 and K = 61. Separate runs for breeds that clustered together shown below (Leroy et al. 2009).

1.3.2. SNPs

More precise information uncovering dog breed clustering was achieved by SNP studies (vonHoldt et al. 2010; Parker et al. 2017; Talenti et al. 2018). Wolves separated first in all studies. This means that no recent dog–wolf admixture was detected (vonHoldt et al. 2010; Larson et al. 2012; Frantz et al. 2016; Parker et al. 2017). In Frantz et al. (2016) study Saarloos wolfdog separated second to wolf with 100 bootstrap support. Similar pattern was found in work done by Mastrangelo et al. (2018). A canine SNP chip of 172 000 SNP markers from LUPA dataset was used (Lequarré et al. 2011) with additional samples of Braque français, type Pyrénées, a French hunting dog. Multidimensional scaling method (MDS) uncovered structure of three clusters. One contained majority of modern European breeds, second group included ancient Asian breeds e.g.: Shar-pei, Siberian Husky, Greenland Sledge Dog, Samoyed, together with Eurasier, Finish spitz, Czechoslovakian wolfdog and Saarloos wolfdog. The third cluster grouped Border terrier, English bulldog, Boxer and English bullterrier. Individuals of wolves and German shepherd were the founding populations of Saarloos wolfdogs and

10

Czechoslovakian wolfdog. Wolfdogs as dog-wolf crosses are genetically shifted from the other dog breeds (Leroy et al. 2009). Vaysse et al. (2011) processed the LUPA dataset using Neighbour-Joining clustering method (Saitou & Nei 1987). Saarloos wolfdog clustered closer to the wolves than Czechoslovakian wolfdogs. This does not correspond with the historical data. Only one female wolf contributed to Saarloos wolfdog's gene fond in 1937 (Saarloos Wolfdog Club UK & Érie 2017). The first Czechoslovakian wolfdogs F1 hybrid was created in 1958 and all together 2 male wolves and 2 female wolves contributed to the gene pool (Smetanová et al. 2015; Caniglia et al. 2018).
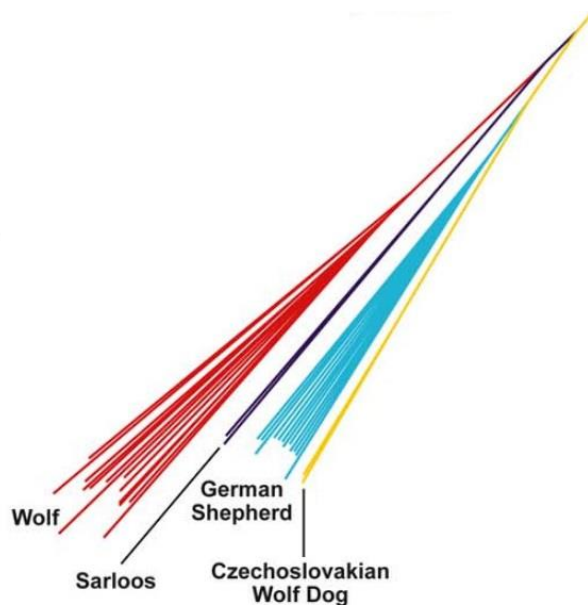
Figure 5: Segment of a Neighbour-joining tree obtained from LUPA project SNP dataset. Saarloos wolfdog clusters closer to wolves than Czechoslovakian wolfdog (Vaysse et al. 2011).

Outside of wolf – dog crossbreeds an African breed Basenji used to separate just after the wolves (vonHoldt et al. 2010; Parker et al. 2017; Talenti et al. 2018). Only in Larson et al. (2012), two Asian breeds Akita and Chow Chow separated before. Basenji is thought to be one of the oldest dog breeds in the world. VonHoldt et al. (2010) even suggests that Basenji have high haplotype sharing with Middle Eastern wolves so there could have been a recent backcross. Next nod separates old Asian breeds like Akita, Chow Chow, Siberian Husky, Alaskan Malamute and Shar-pei. VonHoldt et al. (2010) included two "wild" dogs, Australian Dingo and New Guinea singing dog. These dogs also clustered in branch with the old Asian breeds. We could see underlining of this structure already in Parker et al. (2004). VonHoldt et al. (2010) pointed out that these breeds share

alleles with Chinese wolves. This branching supports claim of two separate domestication events (Frantz et al. 2016).

Another interesting topic is the branching of Nordic Spitz-type breeds. Pohjoismäki et al. (2018) focused deeply in this problematic using SNP markers. In their study, a cladogram of 13 Nordic and Russian breeds was constructed. The data showed cohesion of Finnish Spitz, Norwegian Elkhound and Swedish Vallhund (Figure 5). On the other hand, in Parker et al. (2017) and Talenti et al. (2018) Finnish Spitz clustered on its own (Figure 6). This inconsistency could happen because Parker et al. (2017) collected their data in the United States of America, whereas samples obtained by Pohjoismäki et al. (2018) originated from Scandinavia, a place of origin of most of the breeds.
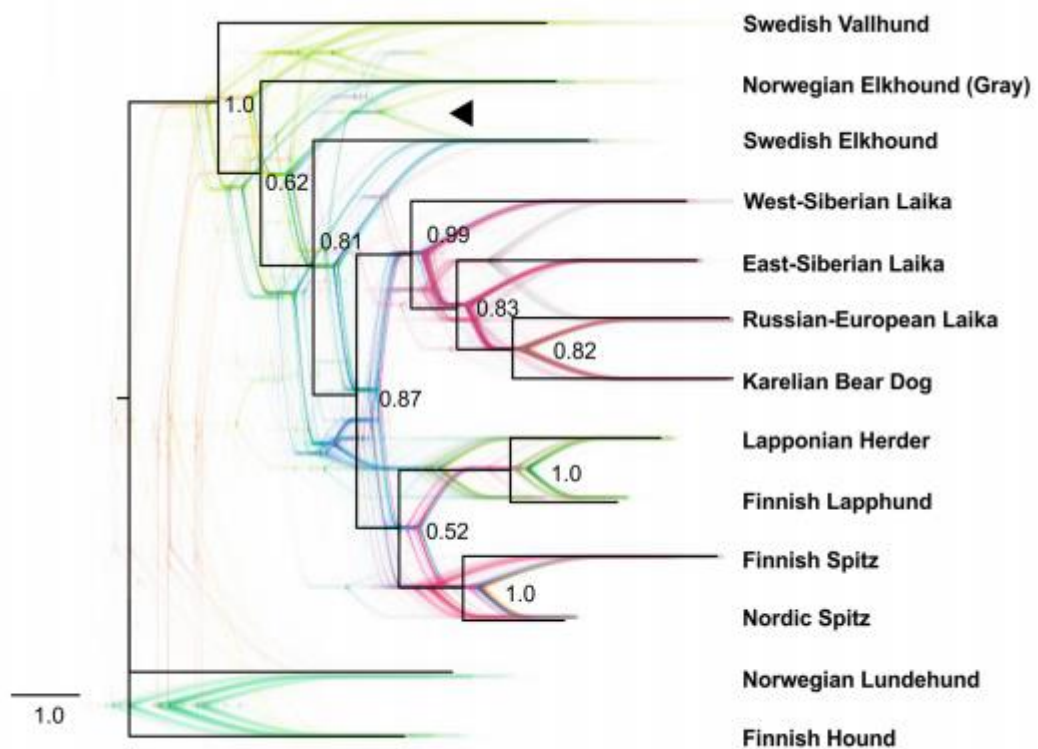
Figure 6: SNP phylogeny of 13 Nordic and Russian breeds acquired using MrBayes (Pohjoismäki et al. 2018).

Two sighthounds, Afghan hound and Saluki, formed a distinct branch in vonHoldt et al. (2010). Similar trend was shown in Larson et al. (2012). These breeds were also considered as ancient. Interestingly, in a study of Parker et al. (2017), Afghan hound and Saluki separated in a distinct "Mediterranean group" together with Great Pyrenees, Ibizan

hound or Pharaoh hound (Figure 6). Parker et al. (2017) also sampled Saluki from the country of origin. These dogs branched monophyletically with the American bred Salukis. This branch contained Ibizan hound and Pharaoh hound, two ancient breeds that were re-created in a recent history (Parker et al. 2007; Boyko et al. 2009). Another level of resolution was brought in by Talenti et al. (2018). This study contained data from Parker et al. (2017) enhanced with several Italian native breeds. With the addition of these regional breeds phylogeny relationships among all breeds changed. For example, the "Mediterranean group" did divide into two parts. First containing Great Pyrenees, Ibizan hound, Pharaoh hound and two Italian breeds, second one consisted of 6 Italian breeds, Caucasian sheepdog, Central Asian sheepdog, Sloughi, Saluki and Afghan hound. While several breeds joined this grouping, Komondor and Kuvasz left the "Mediterranean group" and joined together with Puli and Pumi in the "Hungarian group"(Talenti et al. 2018). Recent SNP studies showed, that phenotype can be sometimes deceiving. According to Fédération cynologique internationale (FCI) all sighthounds fall in the group "X". Results from vonHoldt et al. (2010) and Parker et al. (2017) provide evidence, that sighthound phenotype evolved convergently. With one group being the "Mediterranean" (e.g. Saluki, Azawakh, Afghan hound) and second one named as "UK rural" group (e.g. Greyhound, Whippet, Borzoi) (Figure 6). These two clades show no haplotype sharing (Parker et al. 2017; Talenti et al. 2018). Dog breeds created in the same geographical area show lower level of genetic distances. This underlines the importance of the breeds' place of origin in the whole topic of dog evolution (Frantz et al. 2016, Talenti et al. 2018; Pohjoismäki et al. 2018).

German shepherds are not clustering consistently. In Frantz et al. (2016) German shepherd cluster outside of "modern European breeds". In Parker et al. (2017) it branched paraphyletic to Berger Picard, Xoloitzcuintli (Mexican Hairless Dog), Peruvian hairless dog and Chinook in its own "German shepherd/New world" branch. This unexpected branching was explained by Talenti et al. (2018). The six newly included native Italian breeds fell into the clade with German shepherd. Commonly present western European livestock dog possibly contributed to the formation of this breed in late 19th century. These haplotypes could have been brought to America where they contributed to the creation of the New World dogs. Many European breeds also share German shepherd's haplotypes (vonHoldt et al. 2010; Parker et al. 2017; Talenti et al. 2018). As these six local Italian dog breeds were genetically close to German shepherd, it helped to uncover

the genetic history of New World dogs and German shepherd's part in it. Lupo Italiano was one of the six Italian breeds in this clade. A supposed German shepherd – wolf hybrid which breeding program is controlled by the Italian government. Lupo Italiano appeared to be sister breed with German shepherd, but did not show any significant level of haplotype sharing with the Apennine wolves (Talenti et al. 2018).

There were other differences between the two SNP studies. For example, the first big group of European breed dogs that branched out was the Spaniel-Scent hound clade in vonHoldt et al. (2010), however in Parker et al. (2017) it was the terrier group. Giant Schnauzer branched together with standard Schnauzer in vonHoldt et al. (2010), whereas in Parker et al. (2017) miniature and standard Schnauzer clustered together. Giant Schnauzer was clustered out – alongside Russian Terrier, Rottweiler and Doberman. Glen of Imaal terrier which was interestingly clustering together with mastiff-type dogs (Parker et al. 2007, vonHoldt et al. 2010). Parker et al. (2017) uncovered high haplotype sharing of Glen of Imaal terrier with the mollosiod breeds. This could be the cause of this frequent misclassification.
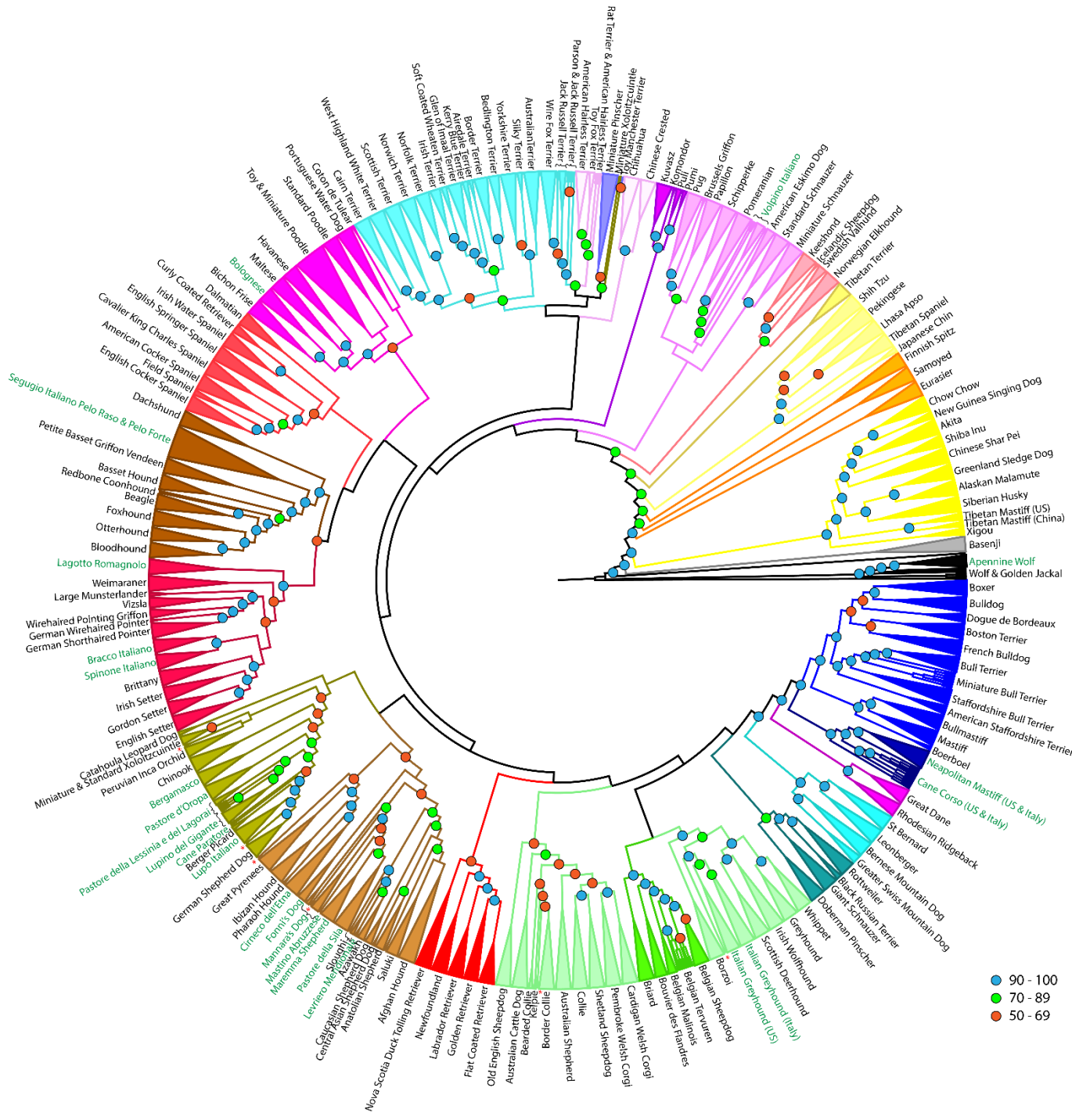
Figure 7: Phylogenetic dendrogram of 182 dog breeds and 16 wild canids based on the genetic distance. Analysis run on 100 bootstraps, nodes supported in > 50 % confidence are figured in colour dots (Parker et al. 2017, Talenti et al. 2018). Groups mentioned in the text: Olive green – German shepherd/New world group; Light brown – Mediterranean group; Mint – UK rural group; Blue – Mastiff group; Turquoise – Terrier group

2. Aims of the Thesis

This thesis serves as a background for further application of the dataset. Therefore, we set following goals:

i) differentiate each breed according to the genotypic data using set of microsatellite loci

ii) test performance of a different clustering approaches

3.   Materials and methods

3.1.   Data collection

The samples of 1238 dogs belonging to 149 breeds (Appendix 1) were gathered on dog exhibitions in central Europe between April 2017 to December 2018(DUO CACIB Bratislava 04/2017; Prague Dog Expo 04/2017 and 11/2017; CACIB Litoměřice 05/2017; Krakow International Dog Show 06/2017; Komárom 3X CACIB 10/2017; Bratislava – National dog show 02/2018; National Dog Show Jelenia Góra 03/2018 and 06/2018; International Dog Show Szilvásvárad 05/2018). The team responsible for sample collection at the dog exhibitions consisted of bachelor students, master students, doctorate students (Ing. Jindřichová, Ing. Neradilová and Ing. Štochlová), head of the laboratory B. Černá Bolfíková PhD. all from Czech University of Life Sciences – Faculty of Tropical Agrisciences, Bc. Ungrová from Charles University and J. Černý PhD. from Biology Center of Czech Academy of Science. The database was completed with the samples obtained directly from the breeders. Dogs were sampled using non-invasive PERFORMAgene PG-100 buccal swabs according to the DNA Genotek protocol PD-PR-099. Non-invasive sampling was used in order to minimize stress of the animals according to welfare. Before sampling the dog's owner was advised not to feed the animal at least 30 minutes prior to the sampling. No dogs were sampled without a permission of the owner. All data had been anonymized. Emphasis was done to sample non related individuals.

3.2.   DNA extraction

DNA was extracted according to the protocol provided by DNAgenotek inc. At the final eluting stage RNA free water was used as a solvent for the DNA. After the extraction DNA concentration and purity was measured using spectrophotometry within λ=240-300nm range at the Nanodrop 2000 spectrophotometer. The extracted DNA was later stored at the temperature of -20°C until further analysis.

3.3.   PCR

Microsatellites STRs analysis loci were chosen to be the cost-effective most appropriate markers which can be later used for follow-up studies and application of this analyses into veterinary-genetic practisefor this work thanks to t. Advantage is their fast mutation rate and wide successful usege in dog population genetics studies (Irion et al. 2003, Koskinen et al. 2003, Parker et al. 2004, Parker et al. 2007, Veit-Kensch et al. 2007, Leroy et al. 2009).

Twenty-one autosomal microsatellite loci of ISAG canine panel were usedselected: AHTk211, CXX279, REN169O18, INU055, REN54P11, INRA21, AHT137, REN169D01, AHTh260, AHTk253, INU005, INU030, FH2848, AHT121, FH2054, REN162C04, AHTh171, REN247M23. AHTh130, REN105L03, REN64E19. In addition to these, 8 other microsatellite markers (FH2096, FH2087, FH2161, FH2140, FH2097, FH2010, vWF, CPH5) were selected from the known literature.

All 29 markers were divided into 3 sets: Primer mix 1, Primer mix 2 and Primer mix 3. Loci were divided into these mixes according to their expected length range, annealing temperature and fluorescent label. PCR was carried out on BIO RAD T-100 thermal cycler, according to the protocol for Type-it QIAGEN microsatellite PCR kit. Conditions for each Primer mix are listed in the Tables 3, 4, 5. Two markers (AHTk121, AHTH130) had to be removed from the further analyses due to the low amplification success.

Step	Temprerature	Time
1.	95°C	5m
2.	95°C	30s
3.	62°C	90s
4.	72°C	30s
5.	go to Step 2.	repeat 28x
6.	60°C	30m
7.	12°C	∞

Table 3: PM1 PCR protocol

Step	Temprerature	Time
1.	95°C	5m
2.	95°C	30s
3.	59°C	90s
4.	72°C	30s
5.	go to Step 2.	repeat 28x
6.	60°C	30m
7.	12°C	∞

Table 4: PM2 PCR protocol.

Step	Temperature	Time
1.	95°C	5m
2.	95°C	30s
3.	63°C	90s
4.	72°C	30s
5.	go to Step 2.	repeat 28x
6.	60°C	30m
7.	12°C	∞

Table 5: PM3 PCR protocol

3.4. Multiplex STR analysis

Before the analysis the Liz500 standard ladder and formamide was added followed by 5 minutes denaturation run at 95°C. After the denaturation samples were taken for multiplex STR analysis on capillary electrophoresis machine Applied Biosystems 3130XL provided by the laboratory of DNA sequencing under Charles University in Prague.

3.5. Data analysis

Data obtained were visualised and edited in the Geneious software and processed via Bayesian MCMC analysis in the STRUCTURE program (Pritchard et al. 2000). Complete dataset was run together at K = 50, 75, 100, 150 with a burn-in period of 200 000 and 200 000 MCMC repetitions. To achieve separation dogs were divided into nine groups according to relationships published in a genomic study by Parker et al. (2017) with additional data by Talenti et al. (2018) (Appendix 1). For each of nine groups separate run with burn-in 50 000 and 50 000 MCMC repetitions was done. Dogs which were not included in the genomic studies we sorted to groups according to their estimated origin. Dog breeds that did not separate clearly in the larger datasets were run with fewer breeds until separation was achieved. STRUCTURE data were compared with Snapclust program utilizing maximum-likelihood genetic clustering using expectation–maximization algorithm on adegenet platform (Jombart 2008; Beugin et al. 2018). Dogs with more than 20 % of missing microsatellite data were excluded from the research. All members of the FTA laboratory team did work on all parts from isolation until preparation for the fragment analysis and data handling.

4.   Results & Discussion

4.1.   Group 1 – primitive, eastern toy and Hungarian breeds

Snapclust correctly assigned 20 breeds out of 29 when K was set to 28. With prior trials we discovered that the separation of Petit Brabancon and Griffon Bruxellois is not possible. Therefore, we set K to (N_{breeds} − 1). In eight breeds (Shiba Inu, Shar Pei, Thai Ridgeback, German Spitz – wolf type, Shih Tzu, Papillon, Slovakian Cuvac, Polish Tatra Sheepdog) one individual always clustered out of the breed. Petit Brabancon and Griffon Bruxellois clustered together as expected. Chihuahuas did not create any distinct cluster, but rather fell into clusters with German Spitz – wolf type (5 individuals), Komondor (3 individuals) and two individuals clustered together with one individual of Shih Tzu and one Shar Pei in their own distinct cluster. According to results of genomic study by Parker et al. (2017) Giant Schnauzers were put apart from Small Schnauzers to group 8.

STRUCTURE at K = 29 assigned 11 breeds into their private clusters with membership coefficient (Q) > 0.85 (Basenji, Siberian Husky, Akita, Samoyed, Shih Tzu, Shipperke, Small Schnauzer, Pug, Chinese Crested Dog, Kuvasz, Komondor) (Figure 8). We chose to test this group on K = 29, because we tested whether Chihuahuas will separate with higher number of possible clusters. STRUCTURE did not express Snapclust's pattern of misclustering. One individual of Shiba failed to cluster within the breed. This might be caused by incorrect sample indication during the sample collection. Shar Pei who clustered together with one Shih Tzu and two Chihuahuas did cluster in STRUCTURE with Siberian Husky at Q = 0.233. The misclustered Shih Tzu did cluster well in STRUCTURE having Q < 0.1 with other breeds. The German Spitz – wolf type that clustered together with German Spitz – miniature did cluster within the breed with Q = 0.020. As this individual shows nothing in common with other German Spitzs – wolf type, it is probably another error of sample origin indication. One individual of Papillon that clustered with German Spitz – wolf type in Snapclust clustered in STRUCTURE the same way with Q = 0.248. No records of Spitz – Continental Spaniel crossings were found. Both breeds cluster on the same branch in Parker et al. (2017). This means that German Spitz and Papillon breeds are genetically close together and therefore allele sharing could occur. Individual of Slovakian Cuvac which clustered as Tibetan Mastiff in

Snapclust, clustered within Slovakian Cuvacs with Q = 0.417 and with Q < 0.1 with other breeds. One Mudi which clustered with Polish Tatra Sheepdogs in Snapclust clustered within the breed in STRUCTURE. Interestingly both clustering programs distinguished Polish Tatra Sheepdog and Slovakian Cuvac. Each breed has almost identical phenotype and breed standards (FCI 2019a, 2019b). Both breeds originated in Tatra mountains separating Slovakia and Poland. It is possible that the country borders and high mountains completely separated the gene pool.
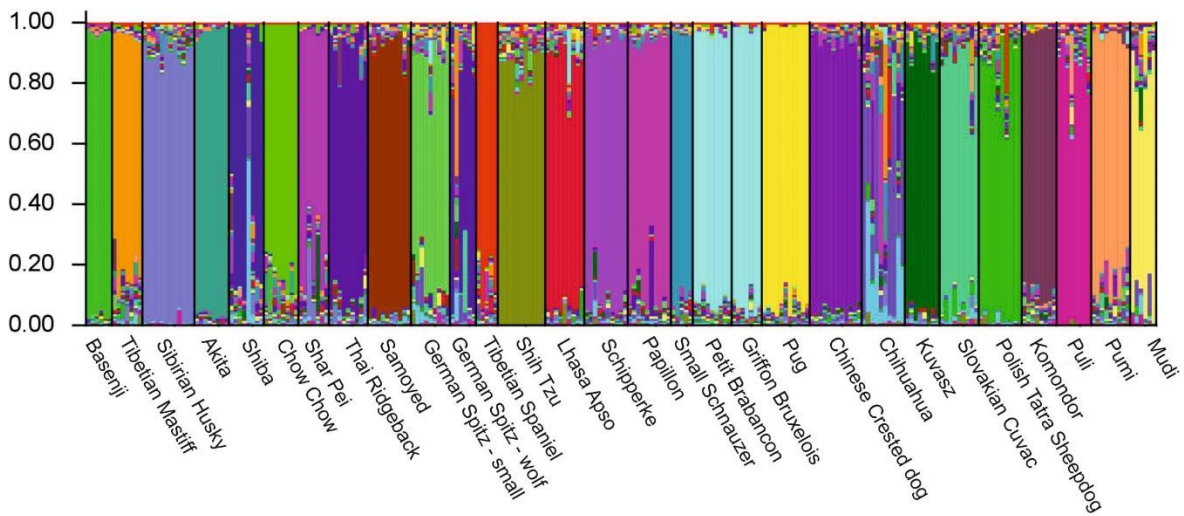
Figure 8: Group 1 - STRUCTURE run of 29 breeds at K = 29. STRUCTURE assigned 26 breeds into their private clusters. Petit Brabancon and Griffon Bruxellois clustered together. Chihuahuas failed to cluster.

A separate STRUCTURE run for Petit Brabnacon, Griffon Bruxellois, Pug, Chinese Crested Dog, Chihuahua and Kuvasz was tried in order to achieve separation of Chihuahuas. At K = 5 every breed created a private cluster with Q > 0.95 except for Petit Brabnacon and Griffon Bruxellois who clustered together in all runs. In this run we were able to separate Chihuahuas. In STRUCTURE runs with higher number of breeds Chihuahuas are not clustering well. Similar result was achieved by Parker et al. (2004). As far as we know, no closer studies were done on population structure of Chihuahuas. More samples of all Chihuahua varieties will be needed to acquire a greater resolution and allele coverage.

A separate STRUCTURE run at K = 2 for Petit Brabancon and Griffon Bruxellois did not find any structure in the samples splitting both breeds in two clusters with Q = 0.5 / 0.5 for Petit Brabancon and Q = 0.497 / 0.503 for Griffon Bruxellois. Petit Brabancon, Griffon Bruxellois and Griffon Belge (not included in our research) are allowed to

interbreed under the rules of Czech Griffon and Brabancon club and every individual born is later assigned to the breed according to its' phenotype (KCHGB 2010). Such a breeding management is producing continuous gene flow between the breeds resulting in impossibility of genetic breed assignment using non-coding markers. However, it is possible to distinguish them based on coat colour and coat type genes (R-spondin-2, Agouti signal peptide gene variations etc.) (Cadieu et al. 2009; Dreger & Schmutz 2011).

Breeds which we characterised as a Group 1, are the closest lineages to the root of the tree in the genomic study (Parker et al. 2017). These breeds are the oldest breeds of modern dogs and therefore their gene pool should have had enough time to develop well defined gene pool. This led us to assumption that these breeds should be well distinguishable based on genotypic data. This group however had the largest dataset and the biggest number of breeds, which tend to decrease values of Q and arbitrary produced "noise" in breed clustering.

4.2. Group 2 - Terriers

In the Group 2 – Snapclust correctly assigned 12 breeds to their own cluster (Russian Toy Terrier, Airedale Terrier, Australian Silky Terrier, Bedlington Terrier, Border Terrier, Cairn Terrier, Czech Terrier, Irish Terrier, Kerry Blue Terrier, German Hunting Terrier, Sky Terrier, West Highland White Terrier). Five breeds had one individual mismatched to a different cluster. One breed had two individuals mismatched. Jack Russel Terrier and Parson Jack Russel tend to cluster together and had many individuals scattered among other clusters i.e. two individuals of each were incorrectly assigned as American Hairless Terrier.

STRUCTURE run of K = 20 was performed for all 20 breeds of the Group 2. results provided higher resolution (Figure 9). Highest Q was achieved for Sky Terriers (0.958). Another seven breeds performed with a Q > 0.9. These were matching with the correctly assigned breeds in Snapclust analysis. The other four breeds which clustered correctly in Snapclust (Russian Toy Terrier, Australian Silky Terrier, Border Terrier, German Hunting Terrier) perform Q > 0.840 except Cairn Terrier with Q = 0.711. However, all individuals of Cairn Terrier expressed the same pattern in STRUCTURE (Figure 9). On average Q = 0.224 fell within the cluster of West Highland White Terriers.

Individual STRUCTURE run for these two breeds at K = 2 achieved complete separation of these two breeds with Q > 0.985. Both breeds share origin in the same area therefore this pattern can be explained by shared ancestry. This explanation is supported by genomic study, where West Highland White Terrier and Cairn Terrier branch as sister breeds (Parker et al. 2017).

Our data showed a shallow structure of individual clustering between Jack Russel Terriers and Parson Jack Russels. On the contrary, in the study of Leroy et al. (2009) at K = 61, both breeds clustered together. Our data therefore show higher resolution power to differentiate between these two breeds. One individual of Jack Russel Terrier clustered as Parson Jack Russel with Q = 0.704. This can be caused either by mistake during sampling or because phenotype standards of the breeds can intermingle and therefore a gene flow between breeds does not have to be completely restricted (Mellanby et al. 2013). One individual of Welsh Terrier did cluster out of the breed completely. In this sample a mistake was made during the process of analysis.
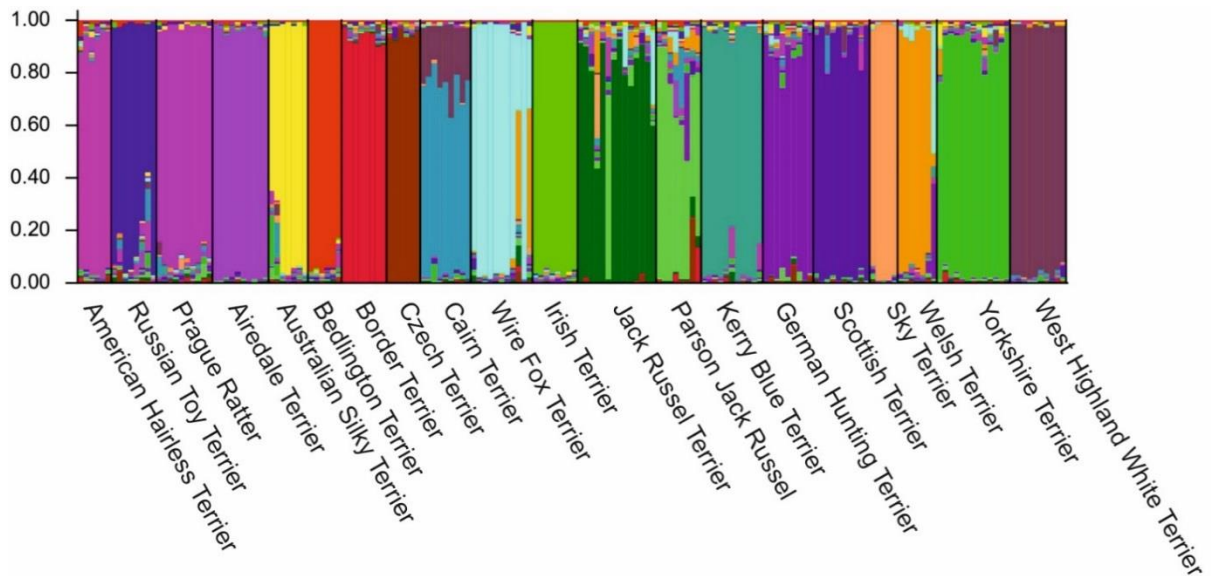
Figure 9: Group 2 – 20 breeds at K = 20 in STRUCTURE. Partial separation of Jack Russel Terriers and allele sharing between Cairn Terriers and West Highland White Terriers.

4.3.   Group 3 – Toy breeds and water Spaniels

In the Group 3, Snapclust outperformed STRUCTURE. At K = 17 Snapclust assigned 15 out of 17 breeds to their own private clusters (Figure 10). Only exception being one individual of Bolognese Dog and one individual of Miniature Poodle. Both clustered as Coton de Tuléar.
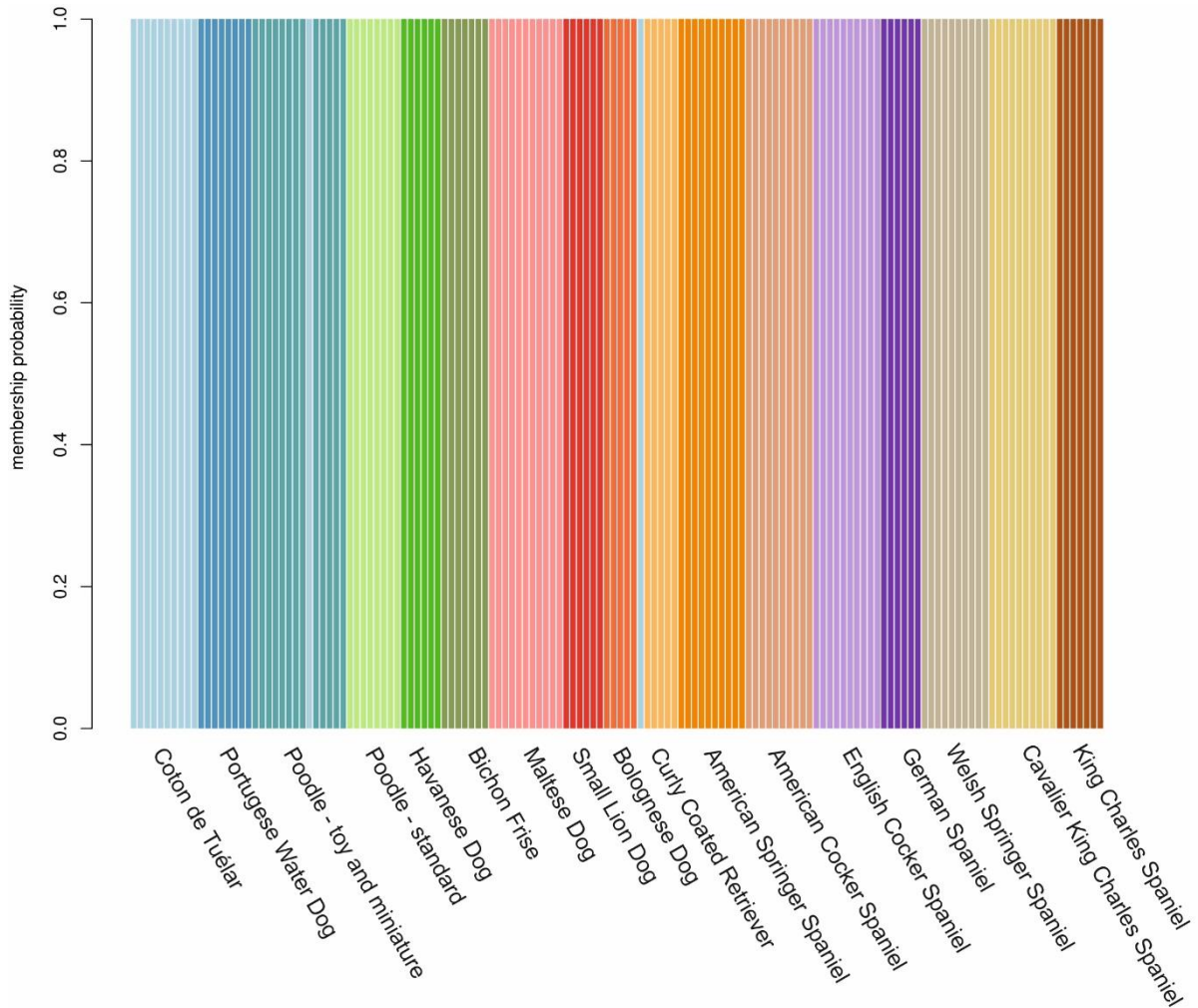
Figure 10: Group 3 - run in Snapclust at K = 17. Every breed assigned into its private cluster. One individual of Miniature Poodle and one Bolognese Dog clustered with Coton de Tuléar.

STRUCTURE at K = 17 assigned correctly only seven breeds to their private clusters with Q > 0.9. However, no dog breeds had their own Q value lower than 0.730 (Bichon Frise). Mean Q value of Bichon Frise was low because one of seven individuals of Bichon Frise expressed tendency to cluster with miniature Poodles (Q = 0.181) and two with Bolognese Dogs (Q = 0.324 and 0.198 respectively). Poodles and Bichons branched together in the genomic study of Parker et al. (2017). Bolognese Dog was not

present in that research, however, Talenti et al. (2018) included this breed. In his study, Bolognese Dogs clustered as a sister breed to Bichon Frise (Talenti et al. 2018). Shared ancestry would be therefore the most probable explanation. The Bolognese Dog individual which mismatched in Snapclust did not express the same way in STRUCTURE (matching the Coton de Tuléar $Q = 0.021$). This individual rather clustered together with one individual of miniature Poodle and formed a private cluster with $Q = 0.616$ for the Bolognese Dog and $Q = 0.720$ for the Poodle. This excessive clustering left two breeds of Cavalier King Charles Spaniel and King Charles spaniel clustering together. Even when runs where $K > 17$ were performed STRUCTURE failed to separate these two breeds. Complete separation (with $Q > 0.96$) was achieved when additional run for these two breeds at $K = 2$ was done. This is pattern corresponds with the previous studies (Leroy et al. 2009).

We were able to differentiate standard Poodles from the other Poodle varieties, which is in agreement with study of Björnerfeldt et al. (2008). Pattern of genetic differentiation between Poodle varieties is visible also when other breeds are included in the analyses, unlike the other closely related breeds (Cavalier King Charles Spaniel & King Charles Spaniel; Collie Rough & Collie Smooth). In Leroy et al. (2009), Poodles clustered together, but there is no closer specification of the sampled variety. In their results, STRUCTURE correctly assigned only 66.7 % of the individuals. In our study Standard Poodles seem to be genetically distant enough from the other Poodle varieties to be correctly assigned as a separate cluster. Björnerfeldt et al. (2008) also clustered Miniature and Toy Poodles according to the coat colour. We were not able to do so. To achieve this separation, we would need many more samples of Poodles. Further analysis of population structure of Poodles in Central and Eastern Europe is advised, to see whether they follow similar pattern of clustering as Poodles in Sweden. Björnerfeldt et al. (2008) states that local regulations and breeding practices can affect population structure. We found similar pattern in Dachshunds as discussed in the following chapter.

## 4.4.	Group 4 – Hunting dogs

In the Group 4, Snapclust clustered correctly 9 breeds out of 19. One individual of Large Münsterländer, one individual of Vizsla and two individuals of Bohemian Wirehaired Pointing Griffon failed to cluster within the breed. German Pointers and Weimaraners clustered separately into clusters according to the coat varieties. In this study we tried to separate Dachshunds into five clusters according to size and coat varieties. Snapclust did not find any structure within the Dachshunds.

At first, STRUCTURE did not separate the breeds correctly due to high variability within the pointing dogs, so we decided to divide the breeds into two subgroups. One contained Dachshunds and Setters, the other one contained Pointing dogs. In the Dachshund – Setter subgroup STRUCTURE at K = 5 correctly assigned all breeds with Q > 0.93. With one more cluster (K = 6), a structure within Dachshunds started to appear (Figure 11). Czech and Slovakian Longhaired Dachshunds separated from the rest. In Czech Republic and Slovakia, it is forbidden to interbreed the coat and size varieties of Dachshunds for more than 40 years. This fact confirms also study of Přibáňová et al. (2009), where observable genetic distances between long coated Dachshunds and other varieties were found. However, in other countries this breeding barrier haven't been set yet or is not strictly respected. According to our results, five individuals of long coated Dachshunds clustered together with other varieties. Non official information admits that some breeders in Poland, Hungary and Russia interbreed these varieties and later separate them according to their phenotype (personal communication with breeders). When K = 7, nine out of twelve individuals of wire coated Dachshunds separated (Figure 11). The other three clustered together with the Smooth coated Dachshund. It is visible that some form of separation already exists. We are just not able to uncover it completely. In few generations it will be possible to separate all coat varieties in the Czech and Slovak Dachshund population. We tried to separate Dachshunds according to the size, but we were not able to distinguish between them.
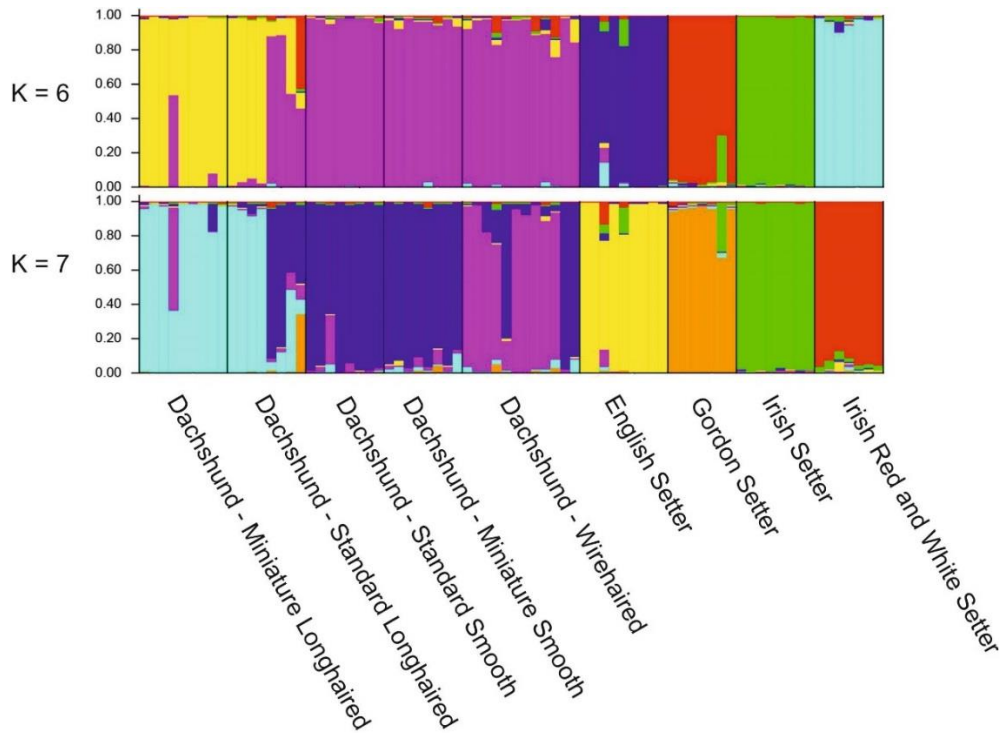
Figure 11: Dachshund and setter subgroup separated at K=6 and K=7. Dachshunds of Czech origin divided from the other Dachshunds at K=6. At K=7, a shallow structure in wirehaired Dachshunds appeared.

In the pointing dogs STRUCTURE clustered six breeds to their own clusters with $Q > 0.85$ (Viszla, Bohemian Wirehaired Pointing Griffon, Bracco Italiano, Czech Spotted Dog, Small Münsterländer, Brittany Dog). The diversity between varieties of German Pointers (Smooth, wirehaired, longhaired) Weimaraners (Smooth, longhaired) were so visible to STRUCTURE, that the program was not able to assign them to their respective breed. Therefore, the Pointing group needed additional sub-grouping. All varieties of German pointing dogs were run separately in one subgroup (German Pointer, Weimaraner) and the rest of the pointing dogs in the other one (Viszla, Bohemian Wirehaired Pointing Griffon, Bracco Italiano, Czech Spotted Dog, Large Münsterländer, Small Münsterländer and Brittany Dog). In the German Pointer group each variety created their own cluster with $Q > 0.713$. Q levels were shared only within the breeds with highest $Q = 0.221$ for Shorthaired Weimaraner clustering to Longhaired Weimaraner (Table 6). To maintain the stability throughout the STRUCTURE runs we would need to have more samples of each variety.

Breed	Q for cluster1	Q for cluster2	Q for cluster 3	Q for cluster 4	Q for cluster 5
VOD	0.005	0.004	0.003	0.979	0.009
VOK	0.046	0.016	0.005	0.221	0.713
NDO	0.946	0.029	0.008	0.010	0.007
DO	0.149	0.106	0.735	0.004	0.006
KO	0.137	0.848	0.005	0.006	0.004

Table 6: Q values at K = 5 of five varieties within 2 dog breeds. VOD - Longhaired Weimaraner; VOK - Shorthaired Weimaraner; NDO - German Wirehaired Pointer; DO - German Longhaired Pointer; KO - German Shorthaired Pointer

The rest of the eight pointing breeds were run separately at K = 8. Structure assigned all breeds with Q > 0.85 to their own clusters, except for English Pointer (Q = 0.82). One individual of Large Münsterländer failed to cluster to its breed. This is a sign of a possible error in the analysis.

4.5. Group 5 – New World and Mediterranean breeds

In the Group 5 – Snapclust correctly assigned 10 out of 14 breeds. In Mexican Hairless Dog and Peruvian Inca Orchid no differentiation was recognised. One individual of Ibizan Hound mismatched to the group of Mexican Hairless Dogs and Peruvian Inca Orchids. Two individuals of Berger Blanc Suisse were incorrectly assigned - one as German shepherd and the other as Czechoslovakian Wolfdog. STRUCTURE separated nine breeds into their private clusters with Q > 0.83 (Louisiana Catahoula Leopard Dog, Czechoslovakian Wolfdog, Hovawart, Pharaoh Hound, Ibizan Hound, Caucasian Shepherd Dog, Afghan Hound, Azawakh and Saluki). In Mexican Hairless Dogs and Peruvian Inca Orchids, STRUCTURE also did not recognise any pattern of differentiation.

German Shepherd, Bohemian Shepherd and Berger Blanc Suisse clustered into their private breed group with Q > 0.86. Clustering of wolfdogs to German shepherd is well supported (Leroy et al. 2009; Vaysse et al. 2011; Smetanová et al. 2015). Additional STRUCTURE run was performed to achieve separation of the shepherd breeds. Two individuals of Saarloos Wolfdog were added. At K = 5 all breeds separated clearly with Q value > 0.88 (Figure 12). These breeds are closely related and come from the gene pool of European shepherd dogs (Talenti et al.2018) and group altogether when the number of

breeds in the STRUCTURE run is high. When the K value rises STRUCTURE tends to reveal other breeds' substructures instead of separating the shepherds. In this case, between breed diversity of the shepherds is lower than within breed diversity of some more ancestral breeds.
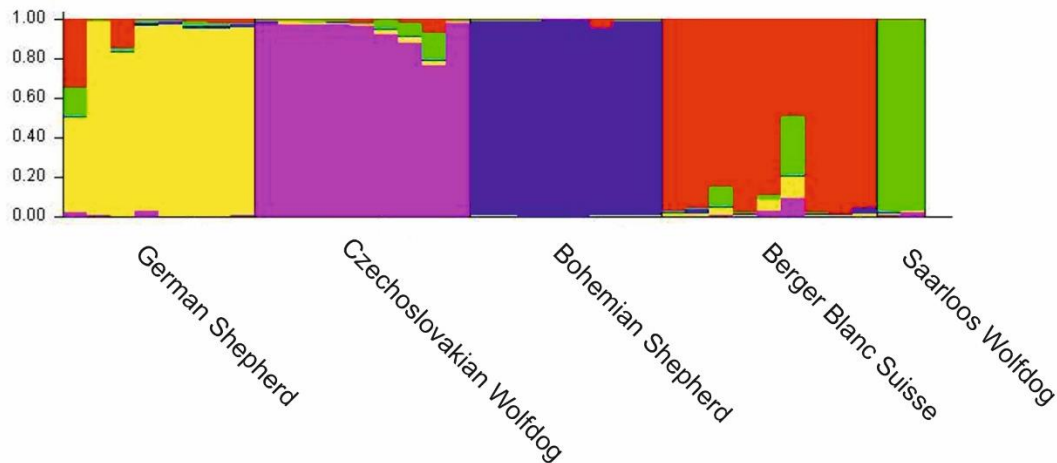
Figure 12: Individual clustering of the closely related Shepherds and Wolfdogs in STRUCTURE at K = 5.

4.6.   Group 6 – Retrievers and Collies

In the Group 6, Snapclust correctly assigned 15 out of 16 breeds. Only one Individual of Border Collie was incorrectly assigned as Australian Shepherd. STRUCTURE run at K = 16 assigned 14 breeds into private clusters with Q > 0.82 (Figure 13). For Nova Scotia Duck Tolling Retriever, Labrador Retriever, Golden Retriever, Flat Coated Retriever, Polish Lowland Sheepdog, Old English Sheepdog and both Welsh Corgis the mean Q value exceeded 0.9. Collie Rough and Collie Smooth clustered together with Q > 0.93. When K < 16 Shetland Sheepdogs clustered inconsistently to another breed's cluster (Collie Rough & Collie Smooth or to Border Collie). In Parker et al. (2004) Collie and Shetland Sheepdog clustered together. To achieve separation in that research, STRUCTURE run of both breeds at K = 2 was necessary. More iterations of STRUCTURE MCMC iterations should be set, to see whether we encountered a coincidence or there is some deeper pattern within this clustering.
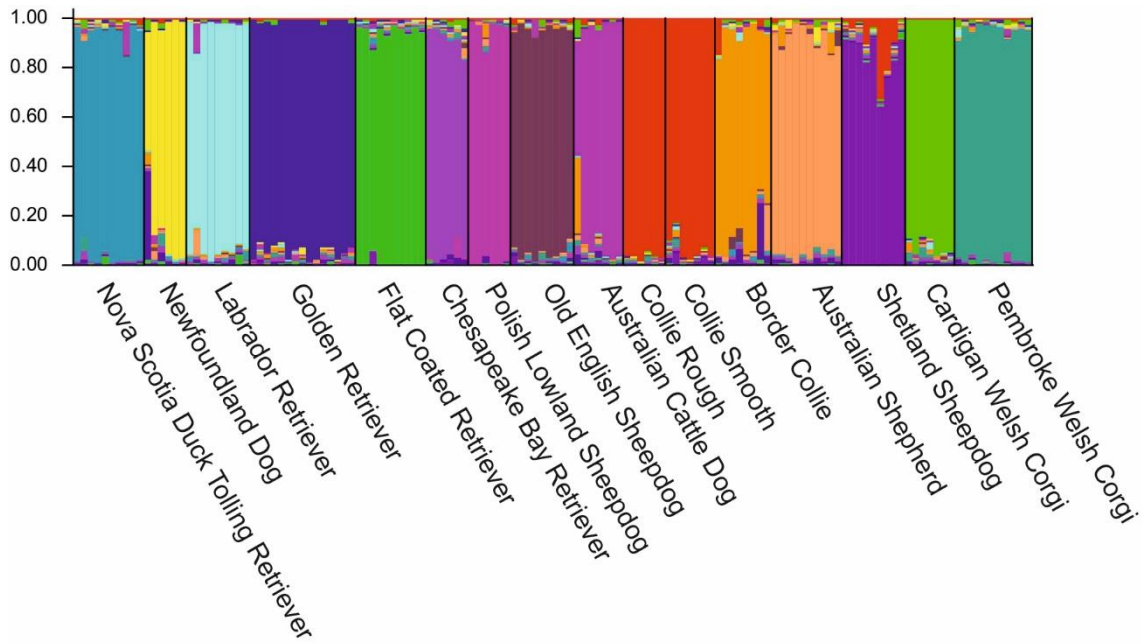
Figure 13: Group 6 – STRUCTURE at K = 16 separated 14 breeds to individual clusters. Collie Rough and Collie Smooth clustered together.

When a separate STRUCTURE run for Collie Rough and Collie Smooth was performed a complete separation was achieved with mean Q = 0.878 for Collie Rough and Q = 0.949 for Collie Smooth respectively (Figure 14). Studies available so far list "Collie" - not specifying what variety is sampled (Parker et al. 2004, 2017; Leroy et al. 2009; vonHoldt et al. 2010). In England, both breeds could be bred together until 1994 (The Kennel Club 2019). Our data indicate that just after 25 years of separate breeding programs we can genetically distinguish between the Collie varieties.
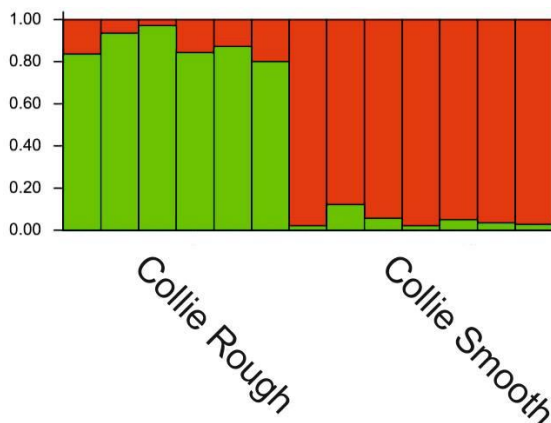
Figure 14: Individual STRUCTURE run at K = 2 for both Collie coat varieties. Both breeds created individual clusters with Q > 0.87.

4.7.　　　　　Group 7 – Continental herder and European Sighthounds

Most fitting result in Snapclust for this group was at K = 13 (N_{breeds} + 1) (Figure 15). When lower K was chosen, individuals of Polish Greyhounds clustered together with Borzois. This effect was probably caused by Dutch Shepherd varieties. There was greater genetic difference measured by clustering programs between long coated and smooth coated Dutch Shepherds than between Polish Greyhounds and Borzois. When K was set to (N_{breeds} + 1) Snapclust clustered all breeds to their own private clusters. Dutch Shepherd was separated into 2 distinct clusters. One containing long coated individuals, the other smooth coated ones.
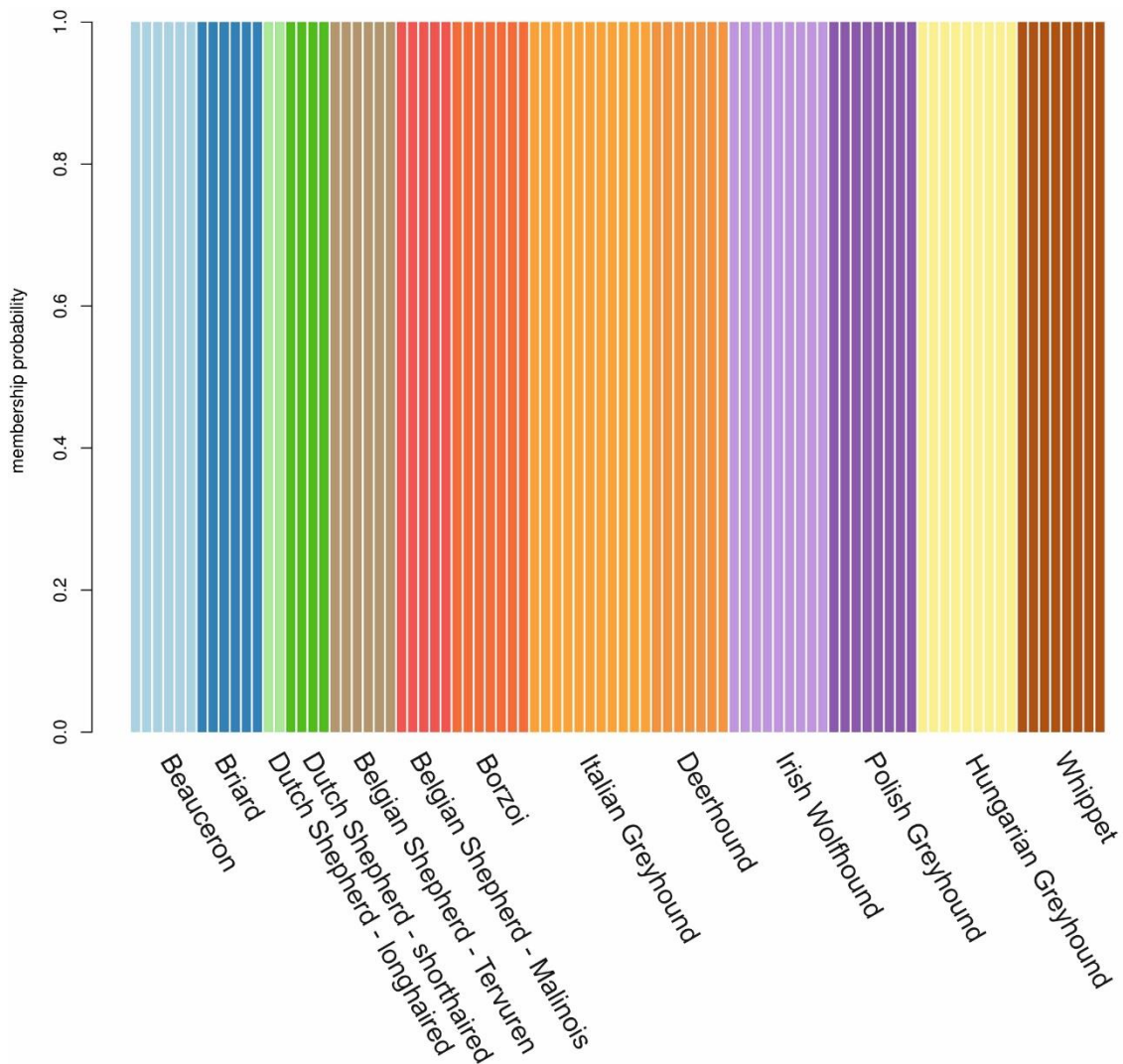
Figure 15: Group 7 - Snapclust run of 13 breeds at K = 14. Eleven breeds created private clusters. Dutch Shepherd was divided to 2 clusters according to its coat variety.

In STRUCTURE, both subgroups (continental herders; European sighthounds) needed to be ran separately on higher resolution (150 000 burn-in / 150 000 repetitions MCMC) because in lower resolutions, STRUCTURE was not able to separate Polish Greyhounds and Borzois but was rather trying to find structure within Hungarian Greyhounds. STRUCTURE run for K = 7 on this higher resolution, clustered all breeds into their private clusters with mean Q > 0.89. We chose to test sighthounds of Middle Eastern origin together with sighthounds of Eastern European origin to further uncover relationships between Borzois and Polish Greyhounds. STRUCTURE run at K = 5 for six breeds tested clustered together Borzois (Q = 0.949) with Polish Greyhounds (Q = 916) (Figure 16). This is interesting, because the breeders believe that Polish Greyhounds originated from Asian sighthounds (Personal communication with breeders). Our data indicate more recent origin and a close relationship with Borzois. Other breeds clustered separately with Q > 0.96.
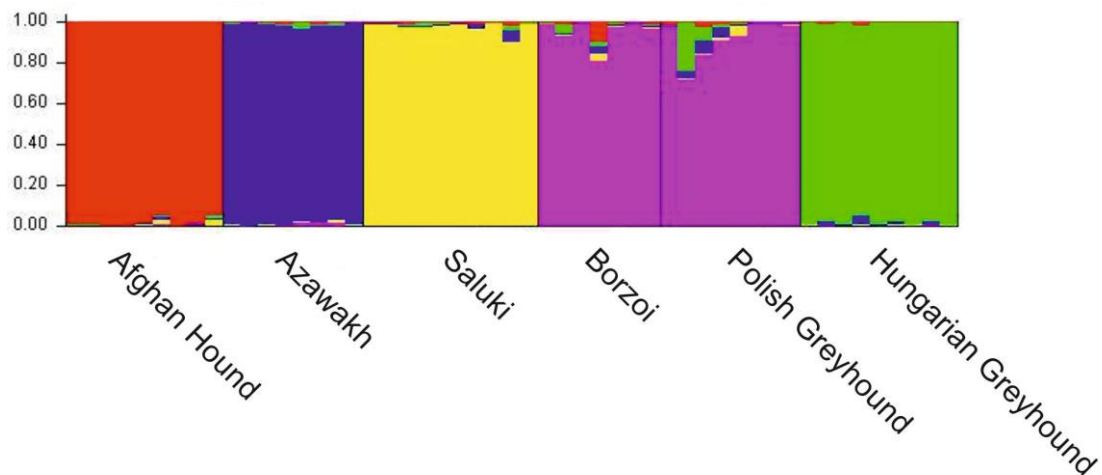
Figure 16STRUCTURE run of three Middle Eastern sighthounds and three Eastern European sighthounds at K = 5. Polish Greyhounds cluster with Borzois in one cluster.

STRUCTURE run at K = 6 was done for all four breeds and two varieties of Dutch Shepherd in continental herder subgroup. In all breeds was Q > 0.93 (Figure 17). Long coated Dutch Shepherds clustered into their own private cluster at Q = 0.988. The rest of Dutch Shepherds segregated into their private cluster with mean Q = 0.858. In Parker et al. (2004) Belgian Shepherd Tervuren and Malinios did not separate even after independent STRUCTURE run at K = 2. We managed to cluster these breeds correctly at K = 6.
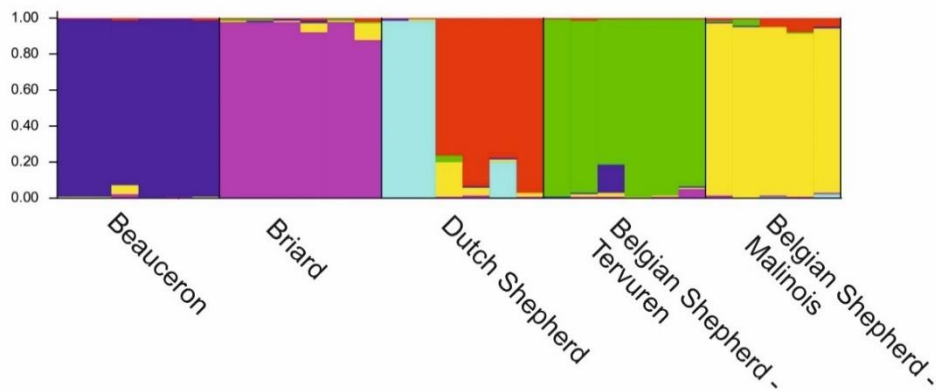
Figure 17: STRUCTURE run of continental shepherd subgroup. At K = 6 a clear separation of longhaired Dutch Shepherds appeared.

4.8.   Group 8 – Drover, Alpine and Great Dane

Snapclust correctly clustered seven out of eight breeds. One individual of Great Dane was incorrectly assigned as Black Russian Terrier. STRUCTURE assigned all breeds within this group into their own clusters with mean Q > 0.91 (Figure 18). Same individual of Great Dane that did cluster with Black Russian Terriers in Snapclust fell within the Black Russian Terrier cluster with Q = 0.171. About 17 breeds formed Black Russian Terriers in the 1950s (AKC 2019a). Great Dane was one of them. This mismatch and allele sharing could be a remnant of these crossings. Even though Doberman, Giant Schnauzer, Black Russian Terrier and Rottweiler are on the same branch in the genomic study by Parker et al. (2017) we managed to separate all breeds to their private clusters.
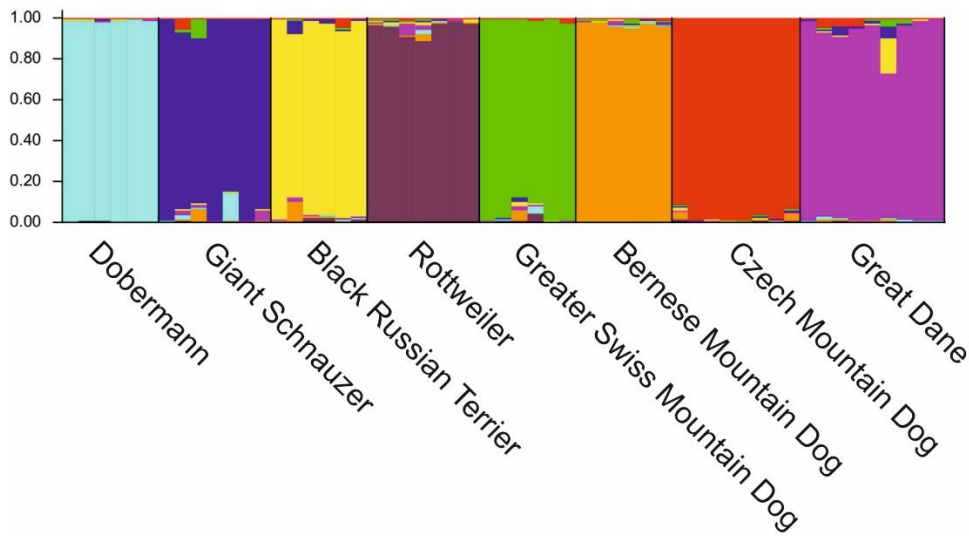
Figure 18: Group 8 - STRUCTURE run at K = 8 separated all eight breeds into their own clusters.

4.9.　　　　　　Group 9 – European Mastiff

In the Group 9 Snapclust correctly assigned 13 breeds out of 14 totals. Only one individual of Cane Corso was inaccurately assigned to English Mastiff cluster. STRUCTURE created individual cluster for each breed at K = 14 with Q value only four times dropping below 0.9 (Bullmastiff, Q = 0.847; Boston Terrier, Q = 0.879; Dogo Argentino, Q = 0.871; American Pit Bull Terrier; Q = 0.88) (Figure 19). Highest mean Q value was achieved for Bull Terrier (0.971). STRUCTURE did not reveal similar mismatching pattern in Cane Corso as was observed in Snapclust, however in 1970s Cane Corso breed went almost extinct (AKC 2019b). To save the breed a possible outcrossing with English Mastiff could have happened. In STRUCTURE, second individual of Bullmastiff fell within cluster of American Bulldog with Q = 0.266. Bullmastiff is a cross between Bulldog and Mastiff with 40 % / 60 % ratio (AKC 2019c). Shared ancestry could be an explanation of this mismatching.
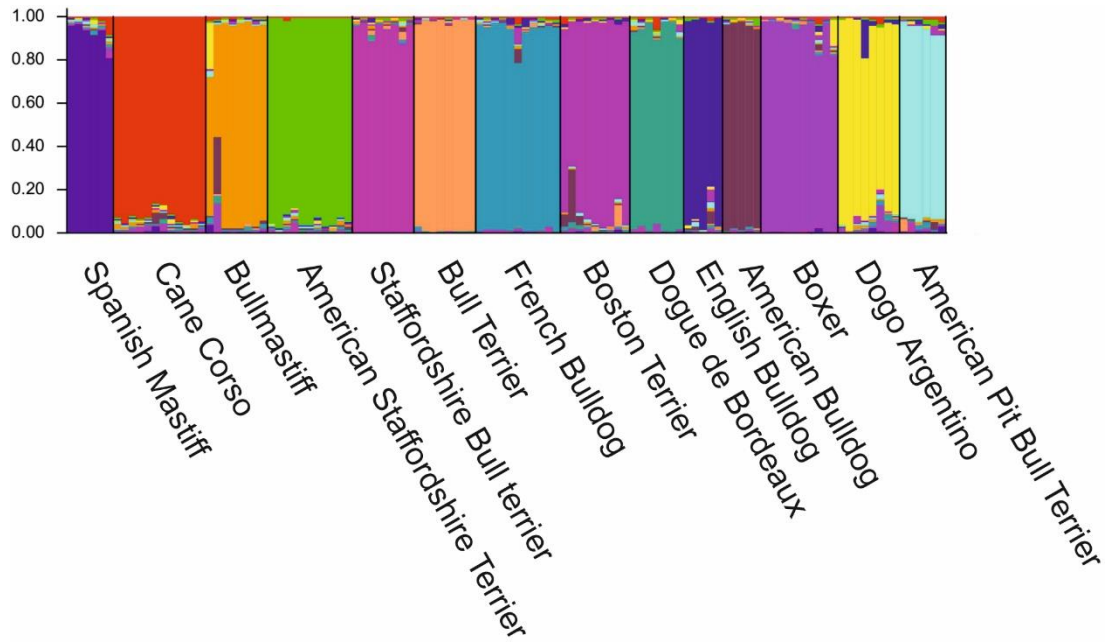
Figure 19: Group 9 - STRUCTURE run at K = 14 separated all 14 breeds into their private clusters.

4.10.   All groups

STRUCTURE runs for all breeds were performed at K = 50, 75, 100 and 150. At K = 50 STRUCTURE correctly assigned 11 breeds with mean Q > 0.85 (Akita, West Highland White Terrier, Scottish Terrier, Sky Terrier, Cavalier King Charles Spaniel, German Shepherd, Afghan Hound, Polish Lowland Sheepdog, Czech Mountain Dog, Bull Terrier, Dogo Argentino). Collie Rough and Collie smooth clustered together in the same cluster with mean Q > 0.85. With higher number of clusters, the mean Q values were dropping. At K = 75, only two breeds exceeded Q > 0.85 (Basenji, Bull Terrier). At K = 100 and 150 only Bull Terrier clustered with Q > 0.85. When number of K was increased, breed assignment was possible, but only based on personal visual evaluation, even though the Q was lower than 0.85. Decreased mean Q values are connected with random assignment of the breeds into other clusters even with the low Q. Considering high tested K, even small assignment (i.e. Q = 0.001) leads to increased "noise" in the bar plot and the mean breed Q. These low individual Q values usually oscillated between 0.001 – 0.005.

We discovered that some varieties of several breeds have bigger genetic distances than expected (i.e. Dutch Shepherd, Collie, Weimaraner, German Pointer). This signal

was visible in STRUCTURE, which tended to create clusters within the breed rather than separating other breeds. The breeds that clustered together in separate runs also clustered the same way in the complete dataset runs. For example, American Cocker Spaniel and English Cocker Spaniel clustered together in K = 50, 75 and 100. STRUCTURE did differentiate these two breeds at K = 150 with Q = 0.588 for American Cocker Spaniel and Q = 0.599 for English Cocker Spaniel. Golden Retriever and Flat Coated Retriever clustered inconsistently. On K = 50 and K = 100 STRUCTURE clustered both breeds together, but on K = 75 and K = 150 both breeds were assigned in their private cluster. This suggest very shallow signal, that would need more repetitions for each K, to see consistency of clustering.

Some breeds clustered together and had to be ran separately (i.e. Petit Brabancon & Griffon Bruxellois; Cavalier King Charles Spaniel & King Charles Spaniel; German Shepherd, Czechoslovakian Wolfdog, Bohemian Shepherd and Berger Blanc Suisse; Collie Rough & Collie Smooth). These breeds tended to stick together also in the big dataset. Separation was achieved only partly for the German Shepherd cluster at K = 150. Bohemian Shepherd (Q = 0.753) and Berger Blanc Suisse (Q = 0.457) separated from a joined cluster of German Shepherd and Czechoslovakian Wolfdog.

Dachshunds clustered together at K = 50 and 75 and separated at K = 100 following the same pattern as in separate runs in Group 4. At K = 150 resolution of this group drastically fell. Because the number of breeds included in the thesis was high and some of them were closely related, we did not manage to cluster the breeds to private clusters as well as other studies (Parker et al 2004, 2007; Leroy et al. 2009).

In this microsatellite analysis, we were not able to separate dog breeds according to their origin, because the modern dogs were often crossed to gain some specific traits in the novel breeds. Using such a small number of markers does not allow us to detect ancestral allele sharing. To achieve this goal a high-resolution SNP analysis would be better for the deeper resolution (vonHoldt et al. 2010; Parker et al. 2017).

STRUCTURE provided greater resolution on larger datasets. Its Bayesian approach is based on allele frequencies. Due to this, STRUCTURE was able to uncover patterns of allele sharing between Cairn Terriers and West Highland White Terriers. Snapclust separated these two breeds into private clusters, because the simulation was based purely on genetic distances. Snapclust outperformed STRUCTURE in processing

speed. Snapclust was more efficient when the dataset was smaller and genetic distances between the breeds were high. When the breeds had closer relationships, STRUCTURE simulations provided greater resolution. In complete dataset of 1238 individuals STRUCTURE provided more detailed results. Runs with higher K decreased Q values. Increase in number of markers could lead to higher resolution and reduction of the "noise" in breed clustering.

We used 27 microsatellites and the number of individuals within a breed ranged from 5 to 14. Parker et al. (2007) used similar number of breeds (132) with four to five individuals per breed but had more markers (96). On the other hand, Leroy et al. (2009) study contained 61 breeds with 20 - 30 individuals per breed on less microsatellites than in our study (21). To amplify the resolution of our study, we recommend either to increase the numbers of individuals per breed especially in the dog breeds that failed to cluster to their own clusters (i.e. Chihuahua, Peruvian Inca Orchid, Mexican Hairless Dog, German Pointer, Weimaraner, Dutch Shepherd). Also, an increase of the number of loci should be considered. STRUCTURE runs for the same K were not repeated, which provides information about the stability of the observed signal. This was due to large amount and complexity of the data connected with high demand on computing power. In the future, more repetitions for every K needs to be done and compared. Single run for each clustering in STRUCTURE only drafts the patterns in the dataset.

5.   Conclusions

We managed to assign breeds to their own clusters when the breeds were sorted in nine groups based on their origin. When the whole dataset was analysed, we were not able to cluster the breeds with high resolution. Dog breeds in Groups 3, 6, 8 and 9 were assigned with higher resolution compared to the other groups. We managed to confidently separate two phenotypically similar breeds which both originate in Tatra mountains (Slovakian Cuvac, Polish Tatra Sheepdog). We were not able to separate Petit Brabancon and Griffon Bruxellois because of rules implied in their breeding program. Cairn Terrier fell into cluster with West Highland White Terrier at $Q = 0.224$, most likely due to shared origin. We managed to uncover a shallow structure between Jack Russel Terriers and

Parson Jack Russels. This structure was not described before. In our study, Standard Poodles separated from other Poodle varieties, this agrees with the known literature. Due to the low number of Poodle samples we did not manage to separate them by coat colour. Separation of some closely related breeds needed to be done at K = 2, which we managed to do for Cavalier King Charles Spaniel with King Charles Spaniel and two varieties of Collies (Collie Rough and Collie Smooth). No microsatellite research done before managed to separate the Collie varieties. Some varieties of the breeds were clustering out of the breeds (German Pointer, Weimaraner, Dutch Shepherd). This finding is not available in the available literature. We uncovered differences in genetic background of Dachshunds bred in Czech Republic and Slovakia compared to Polish and Hungarian dogs. We suggest that it can be an evidence of different breeding policies in the countries. We also found a shallow structure between Czech wirehaired Dachshunds and the other varieties. We managed to clearly separate all breeds within the German shepherd cluster even when Saarloos Wolfdog was added. Our data suggest, that ancestors of Polish Greyhounds are not as ancient as it is believed, but rather more connected to the European Sighthounds. We compared performances of clustering programs Snapclust and STRUCTURE. Snapclust was faster and suitable to use for smaller datasets, where the breeds had greater genetic distances. STRUCTURE performed better in lager datasets and in runs where we aimed to uncover relationships of closely related breeds.

6.   References

AKC – American Kennel Club. 2019a. Black Russian Terrier. AKC. Available from https://www.akc.org/dog-breeds/black-russian-terrier/ (accessed April 2019).

AKC – American Kennel Club. 2019b. Cane Corso. AKC. Available from https://www.akc.org/dog-breeds/cane-corso/ (accessed April 2019).

AKC – American Kennel Club. 2019c.   Bullmastiff. AKC. Available from https://www.akc.org/dog-breeds/bullmastiff/ (accessed April 2019).

BBCA - Biewer Breed Club of America. 2014. Breed History. AKC. Available from http://www.biewerbca.org/breed-history.html (accessed February 2019).

Bennett D, Campbell G, Timm RM. 2016. The dogs of Roman Vindolanda, Part I: Morphometric techniques useful in differentiating domestic and wild canids. Archaeofauna:79–106.

Beugin M-P, Gayet T, Pontier D, Devillard S, Jombart T. 2018. A fast likelihood solution to the genetic clustering problem. Methods in Ecology and Evolution **9**:1006–1016.

Björnerfeldt S, Hailer F, Nord M, Vilà C. 2008. Assortative mating and fragmentation within dog breeds. BMC Evolutionary Biology **8**:28.

Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, Corey L, Degenhardt JD, Auton A, Hedimbi M, Kityo R, Ostrander EA, Schoenebeck J, Todhunter RJ, Jones P, Bustamante CD. 2009. Complex population structure in African village dogs and its implications for inferring dog domestication history. Proceedings of the National Academy of Sciences **106**:13903–13908.

Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, vonHoldt BM, Rhue A, Boyko A, Byers A, Wong A, Mosher DS, Elkahloun AG, Spady TC, André C, Lark KG, Cargill M, Bustamante CD, Wayne RK, Ostrander EA. 2009. Coat Variation in the Domestic Dog Is Governed by Variants in Three Genes. Science **326**:150–153.

Caniglia R. Fabbri E, Hulva P, Černá Bolfíková B, Jindřichová M, Stronen AV, Dykyy I. Camatta A, Carnier P, Randi E, Galverni M. 2018. Wolf outside, dog inside? The genomic make-up of the Czechoslovakian Wolfdog. BMC Genomics **19**:533.

Dreger DL, Schmutz SM. 2011. A SINE Insertion Causes the Black-and-Tan and Saddle Tan Phenotypes in Domestic Dogs. Journal of Heredity **102**:S11–S18.

DDTC - Dandie Dinmont Terrier Club. 2018. History of Kennel Clubs in the UK. DDTC. Available from http://www.ddtc.co.uk/ (accessed February 2019).

FCI - Federation Cynologique Internationale. 2018a. Tatra Shepherd Dog. FCI. Available from http://www.fci.be/en/nomenclature/TATRA-SHEPHERD-DOG-252.html (accessed April 2019).

FCI - Federation Cynologique Internationale. 2018b. Slovakian Chuvach. FCI. Available from http://www.fci.be/en/nomenclature/SLOVAKIAN-CHUVACH-142.html (accessed April 2019).

Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, Linderholm A, Mattiangeli V, Teasdale MD, Dimopoulos EA, Tresset A, Duffraisse M, McCormick F, Bartosiewicz L, Gál E, Nyerges EA, Sablin MV, Bréhard S, Mashkour M. (2016). Genomic and archaeological evidence suggests a dual origin of domestic dogs. Science **352**:1228–1231.

Freedman AH, Gronau I, Schweizer RM, Vecchyo DO, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, Beale H, Ramirez O, Hormozdiari F, Alkan C, Vila C, Squire K, Geffen E, Kusak J, Boyko AR, Parker HG, Lee C, Tadigotla V, Siepel A, Bustamante CD, Harkins TT, Nelson SF, Ostrander EA, Marques-Bonet T, Wayne RK, Novembre J. et al. 2014. Genome Sequencing Highlights the Dynamic Early History of Dogs. PLOS Genetics **10**:e1004016.

Freedman AH, Schweizer RM, Vecchyo DO, Han E, Davis BW, Gronau I, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, Ramirez O, Hormozdiari F, Alkan C, Vila C, Squire K. Geffen E, Kusak J, Boyko AR, Parker HG, Lee C, Tadigotla V, Siepel A, Bustamante CD, Harkins TT, Nelson SF, Marques-Bonet T, Ostrander EA, Wayne RK, Novembre J, Webster MT. 2016. Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. PLOS Genetics **12**:e1005851.

Freedman AH, Wayne RK. 2017. Deciphering the Origin of Dogs: From Fossils to Genomes. Annual Review of Animal Biosciences **5**:281–307.

Germonpré M, Sablin MV, Stevens RE, Hedges REM, Hofreiter M, Stiller M, Després VR. 2009. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. Journal of Archaeological Science **36**:473–490.

Irion DN, Schaffer AL, Famula TR, Hughes SS, Pedersen NC. 2003. Analysis of Genetic Variation in 28 Dog Breed Populations With 100 Microsatellite Markers. Journal of Heredity **94**:81–87.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403–1405.

KCHGB – Klub Chovatelů Grifonků a Brabantíků. 2010. Plemenná kniha. KCHGB. Available from http://www.kchgb.eu/chov/plemena_kniha.htm (accessed April 2019).

Koskinen MT. 2003. Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. Animal Genetics **34**:297–301.

Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, Stahl PW, Piper Pj, Lingaas F, Fredholm M, Cornstock KE, Modiano JF, Schelling C, Agoulnik AI, Leegwater PA, Dobney K, Vigne JD, Vila C, Andersson L, Lindblad-Toh K. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proceedings of the National Academy of Sciences **109**:8878–8883.

Lequarré A-S, Andersson L, André C, Fredholm M, Hitte C, Leeb T, Lohi H, Lindblad-Toh K, 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.

Leroy G, Verrier E, Meriaux JC, Rognon X. 2009. Genetic diversity of dog breeds: between-breed diversity, breed assignation and conservation approaches. Animal Genetics **40**:333–343.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, deJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin C, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli K, Parker HG, Pollinger JP, Searle SMJ, Sutter NB, Thomas R, Webber C, Broad Sequencing Platform members & Lander ES. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature **438**:803–819.

Marsden CD, Vecchyo DO-D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proceedings of the National Academy of Sciences **113**:152–157.

Mastrangelo S, Biscarini F, Tolone M, Auzino B, Ragatzu M, Spaterna A, Ciampolini R. 2018. Genomic characterization of the Braque Français type Pyrénées dog and relationship with other breeds. PLOS ONE **13**:e0208548.

Mellanby RJ, Ogden R, Clements DN, French AT, Gow AG, Powell R, Corcoran B, Schoeman JP, Summers KM. 2013. Population structure and genetic heterogeneity in popular dog breeds in the UK. The Veterinary Journal **196**:92–97.

Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM. 2008. The genomic architecture of segmental duplications and associated copy number variants in dogs. Genome Research **19**:491–499.

Ovodov ND, Crockford SJ, Kuzmin YV, Higham TFG, Hodgins GWL, Plicht J van der. 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.

Pang J-F, Kluetsch C, Zhou XJ, Zhang AB, Luo LY, Angelby H, Ardalan A, Ekstrom C, Skollermo A, Lundeberg J, Matsumura S, Leitner T, Zhang YP, Savolainen P. 2009. mtDNA Data Indicate a Single Origin for Dogs South of Yangtze River, Less Than 16,300 Years Ago, from Numerous Wolves. Molecular Biology and Evolution **26**:2849–2864.

Parker HG. 2012. The History and Relationships of Dog Breeds. In The Genetics of the Dog 2nd Edition. CABI. Cambridge.

Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, Ostrander EA. 2017. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. Cell Reports **19**:697–708.

Parker HG, Kim YK, Sutter BS, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA. 2004. Genetic Structure of the Purebred Domestic Dog. Science **304**:1160–1164.

Parker HG, Kukekova AV, Akey DT, Goldstein O, Kirkness EF, Baysac KC, Mosher DS, Aguirre GD, Acland GM, Ostrander EA. 2007. Breed relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. Genome Research **17**:1562–1571.

Pohjoismäki JLO, Lampi S, Donner J, Anderson H. 2018. Origins and wanderings of the Finnish hunting spitzes. PLOS ONE **13**:e0199992.

Přibáňová M, Horák P, Schröffelová D, Urban T, Bechyňová R, Musilová L. 2009. Analysis of genetic variability in the Czech Dachshund population using microsatellite markers. Journal of Animal Breeding and Genetics **126**:311–318.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics **155**:945–959.

Saarloos Wolfdog Club UK & Érie. 2017. History. Saarloos Wolfdog Club UK. Available from https://www.saarloos.org.uk/welcome/history/ (accessed February 2019).

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

Schelling C, Gaillard C, Dolf G. 2005. Genetic variability of seven dog breeds based on microsatellite markers. Journal of Animal Breeding and Genetics **122**:71–77.

Shannon LM, Boyko RH, Castelhano M, Corey E, Hayward JJ, McLean C, White ME, Abi Said M, Baddley AA, Bondjengo NI, Calero J, Galov A, Hedimbi M, Imam B, Khalap R, Lally D, Masta A, Oliveira KC, Perez L, Randall J, Tam NM, Trujillo-Cornejo FJ, Valeriano C, Sutter NB, Todhunter RJ, Bustamante CD, Boyko AR. 2015. Genetic structure in village dogs reveals a Central Asian domestication origin. Proceedings of the National Academy of Sciences **112**:13639–13644.

Skoglund P, Ersmark E, Palkopoulou E, Dalén L. 2015. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. Current Biology **25**:1515–1519.

Smetanová M, Černá Bolfíková B, Randi E, Caniglia R, Fabbri E, Galaverni M, Kutal M, Hulva P. 2015. From Wolves to Dogs, and Back: Genetic Composition of the Czechoslovakian Wolfdog. PLoS ONE **10:**e0143807.

Talenti A, Dreger DL, Frattini S, Polli M, Stefano M, Harris AC, Liotta L, Cocco R, Hogan AN, Bigi D, Caniglia R, Parker HG, Pagnacco G, Ostrander EA, Crepaldi P. 2018. Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. Ecology and Evolution **8**:2911–2925.

Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, Germonpre MB, Sablin MV, Lopez-Giraldez F, Domingo-Roura X, Napierala H, Uerpmann HP, Loponte DM, Acosta AA, Giemsch L, Schmitz RW, Worthington B, Buikstra JE, Druzhkova A, Graphodatsky AS, Ovodov ND, Wahlberg N, Freedman AH, Schweizer RM, Koepfli KP, Leonard JA, Meyer M, Krause J, Paabo S, Green RE, Wayne RK. 2013. Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. Science **342**:871–874.

The Kennel Club. 2019. Collie Smooth information. The Kennel Club. Available from https://www.thekennelclub.org.uk/services/public/breed/display.aspx?id=5120 (accessed April 2019).

Urfer SR. 2009. Inbreeding and fertility in Irish Wolfhounds in Sweden: 1976 to 2007. Acta Veterinaria Scandinavica 51:21. Available from https://actavetscand.biomedcentral.com/articles/10.1186/1751-0147-51-21 (accessed January 31, 2019).

Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Pielberg GR, Sigurdsson S, Fall T, Seppälä, EH, Hansen MST, Lawley CT, Karlsson EK, Consorium The LUPA, Bannasch D, Vila C, Lohi H, Galibert F, Fredholm M, Hägström J, Hedhammar A. Andre C, Lindbald-Toh K, Hitte C, Webster MT. 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. PLOS Genetics 7:e1002316.

Veit-Kensch CE, Medugorac I, Jedrzejewski W, Bunevich AN, Foerster M. 2007. A heuristic two-dimensional presentation of microsatellite-based data applied to dogs and wolves. Genetics Selection Evolution **39**:447–463.

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE. 1997. Multiple and Ancient Origins of the Domestic Dog. Science **276**:1687–1689.

vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, Reynolds A, Bryc K, Brisbin A, Knowels JC, Mosher DS, Spady TC, Elkahloun A, Geffen E, Pilot M, Jedrezejewski W, Greco C, Randi E, Bannasch D, Wilton A, Shearman J, Musiani M, Cargill M, Jones PG, Qian Z, Huang W, Ding ZL, Zhang YP, Bustamante CD, Ostrander EA, Novembre J, Wayne RK. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature **464**:898–902.

Wang G-D, Zhai W, Yang H-C, Want L, Zhong L, Liu Y-H, Fan R-X, Yin T-T, Zhu C-L, Poyarkov AD, Irwin DM, Hytönen MK, Lohi H, Wu C-I, Savolainen P, Zhang Y-P. 2016. Out of southern East Asia: the natural history of domestic dogs across the world. Cell Research **26**:21–33.

Appendices

List of the Appendices:

Appendix 1: List of breeds by group

Group 1 - Breed name	Code number	Number of Individuals
Basenji	1	6
Tibetian Mastiff	2	7
Sibirian Husky	3	12
Akita	4	8
Shiba	5	8
Chow Chow	6	8
Shar Pei	7	7
Thai Ridgeback	8	9
Samoyed	9	10
German Spitz - small	10	9
German Spitz - wolf	11	6
Tibetian Spaniel	12	5
Shih Tzu	13	11
Lhasa Apso	14	9
Schipperke	15	10
Papillon	16	10
Small Schnauzer	17	5
Petit Brabancon	18	9
Griffon Bruxelois	19	7
Pug	20	11
Chinese Crested dog	21	12
Chihuahua	22	10
Kuvasz	23	8
Slovakian Cuvac (Slovenský Čuvač)	24	9
Polish Tatra Sheepdog (Owczarek Podhalański)	25	10
Komondor	26	8
Puli	27	8
Pumi	28	9
Mudi	29	6

Group 2 - Breed name	Code number	Number of Individuals
American Hairless Terrier	101	6
Russian Toy Terrier	102	8
Prague Ratter (Pražský Krysařík)	103	10
Airedale Terrier	104	10
Australian Silky Terrier	105	7
Bedlington Terrier	106	6
Border Terrier	107	8
Czech Terrier (Český Teriér)	108	6
Cairn Terrier	109	9
Wire Fox Terrier	110	11
Irish Terrier	111	8
Jack Russel Terrier	112	14
Parson Jack Russel	113	8
Kerry Blue Terrier	114	11
German Hunting Terrier (Deutscher Jagdterrier)	115	9
Scottish Terrier	116	10
Sky Terrier	117	5
Welsh Terrier	118	7
Yorkshire Terrier	119	13
West Highland White Terrier	120	10

Group 3 - Breed name	Code number	Number of Individuals
Coton de Tuélar	201	10
Portugese Water Dog	202	8
Poodle - Toy and Miniature	203	14
Poodle – Standard	204	8
Havanese Dog	205	6
Bichon Frise	206	7
Maltese Dog	207	11
Small Lion Dog (Löwchen)	208	6
Bolognese Dog	209	6
Curly Coated Retriever	221	5
American Springer Spaniel	222	10
American Cocker Spaniel	223	10
English Cocker Spaniel	224	10
German Spaniel	225	6
Welsh Springer Spaniel	226	10
Cavalier King Charles Spaniel	227	10
King Charles Spaniel	228	7

Group 4 - Breed name	Code number	Number of Individuals
Dachshund - Miniature, Longhaired	301	9
Dachshund - Standard, Longhaired	302	8
Dachshund - Standard, Smooth	303	8
Dachshund - Miniature, Smooth	304	8
Dachshund - Wirehaired	305	12
English Pointer	331	6
Weimaraner	332	7
German Pointer	333	12
Vizsla	334	14
Bohemian Wirehaired Pointing Griffon (Český Fousek)	335	9
Bracco Italiano	336	7
Czech Spotted Dog (Český Strakatý Pes)	337	9
Large Münsterländer	351	9
Small Münsterländer	352	5
Brittany dog	353	6
English Setter	361	9
Gordon Setter	362	7
Irish Setter	363	8
Irish Red and White Setter	364	7

Group 5 - Breed name	Code number	Number of Individuals
Louisiana Catahoula Leopard dog	401	7
Mexican Hairless Dog (Xoloitzcuintli)	402	10
Peruvian Inca Orchid	403	7
German Shepherd	404	8
Czechoslovakian Wolfdog	405	9
Bohemian Shepherd (Chodský Pes)	406	8
Berger Blanc Suisse	407	9
Hovawart	408	9
Pharaoh Hound	421	5
Ibizan Hound	422	7
Caucasian Shepherd Dog	423	5
Afghan Hound	431	9
Azawakh	432	8
Saluki	433	11

Group 6 - Breed name	Code number	Number of Individuals
Nova Scotia Duck Tolling Retriever	501	10
Newfoundland Dog	502	6
Labrador Retriever	503	9
Golden Retriever	504	15
Flat Coated Retriever	505	10
Chesapeake Bay Retriever	506	6
Polish Lowland Sheepdog (Polski Owczarek Nizinny)	521	6
Old English Sheepdog	522	9
Australian Cattle Dog	523	7
Collie Rough	524	6
Collie Smooth	525	7
Border Collie	526	8
Australian Shepherd	527	10
Shetland Sheepdog	528	9
Cardigan Welsh Corgi	529	7
Pembroke Welsh Corgi	530	11

Group 7 - Breed name	Code number	Number of Individuals
Beauceron	601	6
Briard	602	6
Dutch Shepherd	603	6
Belgian Shepherd - Tervuren	604	6
Belgian Shepherd - Malinois	605	5
Borzoi	621	7
Italian Greyhound	622	11
Deerhound	623	7
Irish Wolfhound	624	9
Chart Polski (Polish Greyhound)	625	8
Magyar Agár (Hungarian Greyhound)	626	9
Whippet	627	8

Group 8 - Breed name	Code number	Number of Individuals
Dobermann	701	6
Giant Schnauzer	702	7
Black Russian Terrier	703	6
Rottweiler	704	7
Greater Swiss Mountain Dog	711	6
Bernese Mountain Dog	712	6
Czech Mountain Dog (Český Horský Pes)	713	8
Great Dane	714	9

Group 9 - Breed name	Code number	Number of Individuals
Spanish Mastiff	801	6
Cane Corso	802	12
Bullmastiff	811	8
American Staffordshire Terrier	812	11
Staffordshire Bull terrier	813	8
Bull Terrier	814	8
French Bulldog	821	11
Boston Terrier	822	9
Dogue de Bordeaux	823	7
English Bulldog	824	5
American Bulldog	825	5
Boxer	826	10
Dogo Argentino	827	8
American Pit Bull Terrier	828	6