

Chinook analysis report

Mars Veterinary

04/1/09

Introduction	2
Data source, error checking, and validation	4
Analysis:	5
Investigating Haplotypes.....	5
Looking at common haplotypes between breeds	6
Looking at diversity based on the quantity of unique haplotypes seen	7
Counting the amount of the same haplotype.....	9
Analysis based on allele frequencies and homozygosity	9
Difference in Allele Frequency between breeds.....	10
Building a Phylogenetic Tree for Chinook data.....	12
Homozygosity in different Chinook groups	14
Analysis based on the Wisdom Panel	16
Breeds detected in Wisdom Panel Professional without the Chinook cluster	18
Creating Alternate Wisdom Panel Clusters.....	19
DIC in the different Chinook groups	20
Discussion:	21

Introduction

This analysis builds on previously communicated reports and figures drawn from Principal Component Analysis plots (PCA) that have been provided to representatives of the Chinook Parent Club and the Chinook Owners Association by Dr Paul Jones.

A request has been received for further analysis to help elucidate how similar the Chinook is genetically in relation to other pure breeds, and how the different groups of the various Chinook breeding population samples provided are related to each other. To do this, we have looked further into the makeup of the breed by looking at where different parts of the Chinook population are similar to other breeds, which would suggest a potential common ancestry at some point in the past.

The Chinook samples provided by the Chinook breed clubs include a set of UKC classified purebred dogs and dogs from the Chinook cross-breeding program.

Since there are a lot of groups, we have attached a nomenclature code related to the group (Table 1) to make it clear which group is being referred to. This is especially important when disambiguating between the UKC classed pure dogs with an open or disputed pedigree (P-P), and a dog from the crossbreeding program which have later become pure (C-P). We provide this table to clarify we have correctly understood the multitude of groups in the corrected data provided.

Identifier	Description	Count
UKC-P	A group of all dogs which are UKC classified as below.	46
U	<i>Chinooks which are UKC classified as purebred dogs and have an undisputed pedigree.</i>	2
T	<i>Chinooks which are UKC classified as purebred dogs yet descended from a UKC classed purebred which was actually a known Chinook cross with a 50/50 German Shepherd Dog / Alaskan Malamute cross. These dogs are also listed as "Cross, Proven Crossbred Ancestor".</i>	8
P-P	<i>Chinooks which are UKC classified as purebred yet with a disputed or open pedigree.</i>	36
CBP	A group of all dogs from the cross-breeding program as listed below. These dogs are back-crosses with purebred Chinooks with a non-chinook ancestor at a given amount of generations back.	95
S	<i>Dogs from the cross-breeding program with a purebred Siberian Husky used in the initial cross.</i>	40
M	<i>Dogs from the cross-breeding program with an unknown Alaskan Malamute mix used in the initial cross.</i>	21
G	<i>Dogs from the cross-breeding program with an initial cross with a dog which was a German Shepherd / Labrador / Siberian Husky mix with at least 50% German Shepherd Dog.</i>	25
FS	<i>Dogs from the crossbreeding program which are crosses between two 5th generation Siberian Husky back-crosses.</i>	3
TS	<i>Dogs from the crossbreeding program which are crosses between a 5th generation Siberian Husky back-cross and a 6th generation Siberian Husky back-cross.</i>	2
FX	<i>Dogs from the crossbreeding program which are crosses between a 5th generation Siberian</i>	4

	<i>Husky back-cross and a 5th generation Alaskan Malamute mix back-cross.</i>	
The sub groups of S, M, and G were also provided with the information of what generation back-cross the dog was. These were:		
2	2 nd generation back-cross.	11
3	3 rd generation back-cross.	15
4	4 th generation back-cross.	30
5	5 th generation back-cross.	21
6	6 th generation back-cross.	9
C	2 nd , 3 rd , and 4 th generation back-crosses which were labeled as C	56
C-P	4 th , 5 th , 6 th , and multiple back-crosses (FS, TS, FX) labeled as P.	42

Table 1: Different sub-groups of dogs in the Chinook data.

There is a difference in pedigree complexity between the pure Siberian Husky back-crosses (S) and the Alaskan Malamute mix (M) and German Shepherd Dog mix (G) dogs. This is because using a mixed dog as a initial cross essentially means that we are looking for purebred-related genetic information at further generations back than the cross itself. This means that the Siberian Husky should be easier to detect than the other crosses used in the cross breeding program, as more Siberian Husky derived genetic material would be admixed into the cross-bred population. This also applies to the UKC classified purebreds (T) descended from the Alaskan Malamute / German Shepherd Dog cross; the genetic information admixed into the Chinook population will be twice as diluted in these dogs. To compensate for the difference of inherited genetic material in these two groups, we could have created a category based on the generation of the breed in the initial cross we were attempting to detect. However, since the proportion of Alaskan Malamute in the Malamute mix (M) is unknown, and the proportions of ancestors in the German Shepherd mix (G) are partially unknown and uneven, this adjustment would have been based on too many assumptions.

A focus of this analysis was to discover what evidence we could find regarding breeds that could have been present in the Chinook founding population; some of these have been suggested to be Siberian Huskies, Alaskan Malamutes, German Shepherd Dogs, and Labrador Retrievers. In addition, attention has been paid to the Belgian herding breeds such as the Belgian Sheepdog, Belgian Tervuren, and Belgian Malinois, and the Gray Wolf as they were thought to have been possible contributors to the purebred Chinook breed development. As one of the original Chinooks was described to be a Mastiff / Husky cross, it would be especially interesting to see if any mastiff breeds appear. The data analysis performed in this report, necessarily limits the breeds examined in some analysis depending on the question studied, so not all were used routinely. This is noted in the appropriate analysis section.

Previous analysis of the Chinook centered around comparing the basic breeds identified in the output Wisdom Panel MX standard maximum likelihood result with the given ancestry and using Principle Component Analysis to plot different combinations of breeds and groups to see if the

groups matched the clusters seen in the PCA analysis. Since this data was communicated, we have been extending the methods with which we use to analyze populations. We have used haplotype analysis from the PHASE software to find genetic differences between closely related terrier breeds, an investigation which led to a region of the genome related to white spotting being the key driver of differentiation between the groups. We have also used haplotype analysis to investigate diversity within another terrier, a reasonably rare breed with small effective population size. Other analysis, such as the Neighbour-Joining tree analysis has also been useful in developing our understanding of the relationship between closely related breeds and higher order breed groups.

The algorithm used for the previous analysis has recently had an update with the amount of breeds it can distinguish, and it has been enhanced based on our increasing knowledge of pure-breed homozygosity patterns, which allows us to increase our accuracy in detecting purebred dogs. Although we have performed analysis which suggests that results of dogs without these added breeds do not change, we hope this will increase the relevancy of the analysis for the Chinook breed clubs.

Data source, error checking, and validation

Constant error checking and validation are major factors in any project or study involving large quantities of data and possible categories, and great care was used to verify analysis.

Samples provided by the Chinook breed clubs were previously genotyped using the Wisdom Panel which gives a value for every Single Nucleotide Polymorphism (SNP) in the panel, which can be classified into four states; missing, homozygous for the first allele, heterozygous, or homozygous for the second allele. This is the data format used in all analysis.

At least 97% of the Chinook samples had data for more than 280 SNPs present, with a single sample having 226 SNPs. Although missing data compromises the accuracy of some analysis, steps were taken to handle missing data throughout the analysis.

This analysis continued from previous work with feedback which observed some inaccuracies in the data. These were fixed before the analysis continued. This includes modifications to samples 19, 27, 60, 79, and 81. 74 was corrected to 6s due to the code in the pedigree info.xls sheet. Also, due to the code in the genotype file being 074-4G instead of 074P-6S, parsing of the code failed and this dog was not included as being in the 6 generations back group.

Analysis:

This section identifies the different analyses used, from investigating haplotypes to looking at analysis based on allele frequencies and homozygosity, to analysis based on the Wisdom MX algorithm itself including using different clusters, no clusters, and DIC.

Investigating Haplotypes

In every dog, for every chromosome, there are two chromatids, and one is inherited from each parent for a given cross. Knowing what data is descended from which chromatid allows us to map which dog breeds are more likely to have been used in the ancestry of a single dog. In a back-cross, we would expect to see some shared chromatids between the back-cross breed and the admixed breed(s) used in the initial cross. Dogs only transfer a single chromatid to their offspring, and there is a 50/50 chance which chromatid will be transferred from that parent. In the case of a first generation back-cross, this means that the genetic material from the initial cross is halved at each subsequent generation. We can predict how many chromosomes on average would be inherited in the n^{th} generation cross, though this is a random process with sometimes more or less chromosomes inherited by chance due to the laws of mendelian inheritance. Table 2 shows the data for the first few generations, with the average proportion of genetic material inherited by descent from the initial dog used in the back-crosses can be calculated from $1 / 2^{x-1}$, where x is generation number. This can be seen to drop off quite rapidly as generations increase. Since the Wisdom Panel focuses on 25 out of the 38 autosomal (non-sex) chromosomes for breed detection in a mixed breed dog, this suggests we might only have a single chromosome or less to detect from the initial cross by the fifth and sixth generations of backcrossing.

Note that a process called recombination usually serves to shuffle the chromosome genetic material, however in the Wisdom Panel we use closely clustered marker windows in order to minimize and effectively negate the effect of recombination in defining recent heritage patterns.

Generation	Average proportion of chromosomes inherited.	Predicted chromosomes (out of 38)	Predicted chromosomes (out of 25)
1	1/1	38	25
2	1/2	19	12.5
3	1/4	9.5	6.25
4	1/8	4.75	3.13
5	1/16	2.38	1.56
6	1/32	1.19	0.78
7	1/64	0.59	0.38

Table 2: Predicted amount of haplotypes inherited containing genetic information from the original sire and dam.

Although we are interested in the data from each chromatid, we get the information from the Wisdom Panel for two chromatids at a time as a dog is diploid. We can use some software called PHASE to infer what the data was on each chromatid, given the input data. We refer to this data on each chromatid as a haplotype. Since these haplotypes were inferred, they are not necessarily perfectly accurate, but they are the most likely estimate possible from population derived data.

When we run the data through PHASE, we calculated the haplotypes independently for each specific sub-group in an analysis. If we did not have this independence, some haplotypes would be predicted as more likely in all samples.

One of the issues with this kind of analysis occurs when asking PHASE to calculate the haplotypes with missing data. To minimize this, for this analysis we removed SNPs with the most missing data for the samples used.

Due to the diversity of the Labrador Retriever breed, we decided to remove it from the haplotype analysis. Although we did start this analysis with the breed, the large quantity of haplotypes discovered in the breed overwhelmed the analysis, and made a visual analysis cumbersome.

We used this analysis to see if haplotypes were conserved between the initial cross and the back-cross in the crossbreeding program, and to see if haplotypes were conserved between the different Chinook groups, yet not in the other breeds. We also used them to explore the diversity of the Chinook compared to other breeds, and investigated the likelihood of offspring having two haplotypes which were the same (homozygous) in different groups.

Looking at common haplotypes between breeds

The main part of the haplotype analysis was designed to examine the haplotype data for each chromosome by eye to see if there were any striking patterns we would expect to see in the crossbreeding program and the Chinooks as a whole:

- Cases where the initial cross breed heritage was conserved in the back-crosses. Since we did not have the actual haplotypes of the initial dogs used in the crosses, this was simulated from haplotypes observed in purebred dogs from the Siberian Husky, Alaskan Malamute, and German Shepherd Dog breeds.
- Cases where the predicted haplotypes found within the breed were uniquely identified in the Chinook, and not in the other breeds analyzed.

The data in Table 3 has thick boxes around clusters which show there are chromosomal haplotypes which are shared between the back-cross breed and the breed in the initial cross which were not observed in the pure Chinook population. This shows that although we have predicted that these haplotypes should appear rarely, they do appear in the crossbred dogs.

	0001	0010	0101	0110	1001	1010	1101	1110
Alaskan Malamute	1	38			6	21		
Chinook_Cross (C, T)	10	9			7	71	21	
Chinook_Possible (P-P)	1				3	53	15	
Chinook_Pure (C-P, U)					13	64	11	
Chinook_f (FS, FX)					3	9	4	
Chinook_g (G)	10					33	11	
Chinook_m (M)		9				28	5	
Chinook_s (S)					15	50	9	
Chinook_t (TS)					2	11	3	
German Shepherd Dog	70	11			50	29	15	23
Siberian Husky	56	29	1	3	3	34	6	

Table 3: Inferred haplotypes shared between groups on chromosome 25.

Table 4 shows instances where haplotypes appear to be conserved within the different Chinook groups, yet are found at low frequency or are not observed in the other breeds. These unique haplotypes suggest that either the haplotype came from a breed not examined here, or that this haplotype has developed independently in the Chinook breed.

	0000	0001	0010	0011	0100	0101	0110	0111	1001	1010	1011	1100	1101	1111
Alaskan Malamute	3	27		24		9	1	1			1			
Chinook_Cross (C, T)	1	10	3	44		27	3	11		6			4	9
Chinook_Possible (P-P)		13		34	6	6		2				2		9
Chinook_Pure (C-P, U)		12		43		11	5	10		4	3			
Chinook_f (FS, FX)			2	6		5				2	1			
Chinook_g (G)		6		14		18		5					1	10
Chinook_m (M)		14	2	2		7	3	11			3			
Chinook_s (S)		2		44		5	3	11		8	1			
Chinook_t (TS)	2	2		10				2						
German Shepherd Dog	18	31		1	14	89						14	30	1
Siberian Husky		26	6	50		7	1	40	1		1			

Table 4: Inferred haplotypes shared between Chinooks on chromosome 6.

Looking at diversity based on the quantity of unique haplotypes seen

Using haplotype analysis, we can examine the diversity of populations based on the variety of haplotypes found. Table 5 shows that that the dogs from the cross-breeding program and pures with a known cross (C, T) are more diverse haplotypically than the groups with pure dogs. It can be seen that there is an advantage of adding the dogs classed as purebred in the cross-breeding program (C-P) dogs to the UKC population (P-P, U), since there would be 54 more haplotypes in the population, which would be 27% more. Although a portion of these haplotypes will have been inherited from the

initial cross used in the back-cross, Table 2 suggests this would not be that many. Possible explanations for this could be due to a degree of error in the PHASE imputation, or that some of these haplotypes are ancestral Chinook haplotypes that have been lost by genetic drift or potentially modified through mutation or recombination events in the purebred population, yet persist in the cross-breeding program.

Group	Haplotypes seen in group (for 25 chromosomes analyzed)
All Chinooks	371
Pure Chinooks and Pures of Crossbred Ancestry (P-P, C-P, U)	256
Crossbred Chinooks (C, T)	288
Possibly Pure Chinooks (P-P, U)	202
Purebred Chinooks of Crossbred Ancestry (C-P)	191
Chinook_f (FS, FX)	109
Chinook_g (G)	166
Chinook_m (M)	171
Chinook_s (S)	194
Chinook_t (TS)	122
German Shepherd Dog	649
Alaskan Malamute	479
Siberian Husky	598

Table 5: Amount of haplotypes detected in the different groups.

We can also make a comparison between the Chinook and other breeds, showing that the Chinook has less haplotype variation than other breeds. The amount of samples used for this analysis can be seen in Table 6, which shows that even though the Alaskan Malamute had only 33 samples (approximately the same as the Chinook: Pure (possible) category, it had more than twice the number of haplotypes detected. It is likely that such an effect is linked to the size of and diversity present in the effective breeding population for each breed, and how related the dogs were from

the individual samples taken. The reference Alaskan Malamute samples originated from different locations across America and Europe.

Breed	#Samples
Alaskan Malamute	33
Chinook	139
German Shepherd Dog	99
Siberian Husky	66

Table 6: Count of samples used for haplotype analysis.

Counting the amount of the same haplotype

As part of our studies into dog diversity, we have developed a metric which looks at the shared haplotypes between two dogs for a single chromosome, and predicts how many of the four possible ancestor combinations would yield two chromatids with the same haplotype in a cross. This metric gives an indication of which combinations of dogs would produce the most variation in the population if selected for breeding. Table 7 shows this metric for the different categories, with a lower number being more diverse. This shows that although there is a lot of diversity in the dogs from the cross-breeding program (C), the pures from this program (C-P) are less diverse than the Possible Pures (P-P). Even so, when we combine the pures from the crossbreeding program (C-P) with the Possible Pures (P-P), we get a gain in diversity, and reduce the likelihood that we will get two haplotypes the same in the offspring of this group.

	Cross	Possible	Pure
Crossbred (C)	27.3%	28.1%	28.5%
Possible Pures (P-P)	28.1%	31.5%	30.4%
Crossbred Pures (C-P)	28.5%	30.4%	32.7%

Table 7: Average Diversity Metric for different Chinook categories.

Analysis based on allele frequencies and observed homozygosity

For a single SNP, allele (variant) frequency and allele homozygosity are similar measures in that allele frequency shows the population frequency of an allele at the SNP locus, and allele homozygosity shows the frequency of this allele being homozygous in the population studied (since there are two chromatids). If the breed has lots of variation, the homozygosity can be quite low, and the allele frequency will be closer to 0.5, since either allele is equally likely. If the breed has less variation, the homozygosity tends to be higher, and the allele frequency will be closer to 0 or 1. Both measures are generally averaged over a group of SNPs on a number of dogs to be informative for a given dog or population.

We have noticed that average allele frequencies will often be similar in genetically related breeds, this allows us to use the difference in allele frequencies to estimate how dissimilar breeds are. We have used this data both individually and together to construct a Neighbor-Joining tree.

In previous work, we have noticed that purebred dogs generally have an average homozygosity value of over 0.59 at the markers on the Wisdom genotyping panel, and mixed bred dogs generally have a homozygosity value below this. We have seen in synthetic dogs that in a theoretical back-cross population, the higher the average homozygosity will be of the population, though we have not had a real-life, documented example to confirm this until the Chinook back-crosses.

Difference in Allele Frequency between breeds

To find the closest breeds in the Wisdom Panel to the different Chinook Clusters, we took the average allele frequency for each breed and found the minimum average difference in allele frequency between these. We then listed the closest breeds which have the lowest average difference in allele frequency. This method does have a couple of weaknesses.

- The most heterozygous breeds will have lower distance and come up more often in the results. This is because a single SNP of two homozygous breeds may have a distance of allele frequency of 1, yet a heterozygous breed has a maximum difference of .5 to any other state.
- This analysis may ignore information on a chromosome if the haplotypes are very different in that breed.
- This analysis is less likely to work as well on groups with few samples (sampling error).

To remove outliers, we first found a list of the most heterozygous breeds in our dataset. These are shown in Table 8.

Rank	Breed
1	Anatolian Shepherd Dog
2	Plott
3	Parson Russell Terrier
4	Italian Spinone
5	Lancashire Heeler
6	Chesapeake Bay Retriever
7	Fox Terrier (Toy)
8	Belgian Malinois
9	German Wirehaired Pointer
10	Silky Terrier
11	American Eskimo Dog
12	German Spitz

Table 8: Most heterozygous breeds in our breed datasets.

Table 9 shows the closest breeds that were found to the different Chinook groups using this methodology. For most of these, the closest breed (and hence the largest proportion of shared genetic material) was the group of dogs in the Possible Pure group (P-P). The closest few breeds detected were also various Chinook groups, showing that the Chinook is the closest breed to all of these clusters, and not a different breed. After removing the most heterozygous breeds from the ranking (grayed out), there was a large contingent of mainly herding breeds in the analysis, which was found across all groups. Surprisingly, we did not see the breeds which have been used in the cross-breeding program, which makes us believe that most signal strength was coming from the

pure Chinook part of the DNA. We found it unusual that both the Bouvier Des Flanders and the Giant Schnauzer came up in this analysis; the similarity to the two breeds makes it more likely that they are not a false positive. It should also be noted that the Bouvier is a sturdy herding breed from Flanders, a region covering parts of modern-day Belgium, France, and the Netherlands. Two northern Spitz-type dogs, the Malamute and the Norwegian Elkhound also appeared in the analysis, which is fascinating, since the Norwegian Elkhound is a breed with a long-standing sledding history in its native Norway. Although it is possible the German Spitz was a relevant result, it is ranked #12 in the most heterozygous breeds, so was discounted.

Chinook: Crossbred C (C)	Chinook: Crossbred P (C-P)	Chinook: Pure P (P-P)	Chinook: Pure U (U)	Chinook: Pure T (T)
Chinook: Pure P	Chinook: Pure P	Chinook: Crossbred C	Chinook: Pure P	Chinook: Pure P
Chinook: Crossbred P	Chinook: Crossbred C	Chinook: Crossbred P	Chinook: Crossbred P	Chinook: Crossbred P
Chinook: Pure T	Chinook: Pure T	Chinook: Pure T	Chinook: Pure T	Chinook: Crossbred C
Chinook: Pure U	Chinook: Pure U	Chinook: Pure U	Chinook: Crossbred C	Chinook: Pure U
Australian Shepherd	German Spitz	German Spitz	Alaskan Malamute	German Shepherd Dog
German Spitz	Australian Cattle Dog	Australian Shepherd	Belgian Tervuren	Australian Cattle Dog
Russell Terrier	Border Collie	Australian Cattle Dog	Bouvier Des Flanders	German Spitz
Plott	Australian Shepherd	Border Collie	Belgian Malinois	Belgian Tervuren
Australian Cattle Dog	Bouvier Des Flanders	German Shepherd Dog	German Shepherd Dog	Australian Shepherd
Anatolian Shepherd Dog	Russell Terrier	Russell Terrier	Schnauzer (Giant)	Chesapeake Bay Retriever
Border Collie	Plott	Schnauzer (Giant)	Golden Retriever	Plott
Schnauzer (Giant)	Anatolian Shepherd Dog	Plott	Chesapeake Bay Retriever	Border Collie
Dachshund	Dachshund	Belgian Tervuren	Icelandic Sheepdog	Chihuahua
Belgian Tervuren	Basset Hound	Norwegian Elkhound	Plott	Belgian Malinois

Table 9: Closest Breeds to five different Chinook subsections. Herding breeds are in green.

Building a Phylogenetic Tree for Chinook data

Phylogenetic trees are a visualization of how species have evolved over time from diverging from an ancestral population. We can use a matrix of differences in allele frequency as a distance matrix to create a phylogenetic tree using the Neighbour-Joining tree generation method.

Restrictions on the amount of data used to create the tree and the tendency of heterozygous breeds can cause errors in this approach. The method is also not very good for showing instances where breeds are formed from multiple combinations of breeds. However, the Neighbor-Joining method has been proved useful to find a tree which is close to the real phylogenetic tree particularly when the relationship is strong.

To build the trees, we created a distance matrix of allele frequencies and put this through the Neighbour-Joining algorithm. We looked at some different groups:

1. The Chinook against all the breeds, this was created by using an average allele frequency for each breed.
2. The different Chinook groups against all the breeds. This was created by using an average allele frequency for each breed.
3. The different Chinook samples against the samples in the Siberian Husky, Alaskan Malamute, and German Shepherd Dog.
4. The different Chinook samples against themselves.

The Chinook in relation to all breeds:

The Chinook itself was classified next to the German Shepherd Dog, Leonberger, and Norwegian Elkhound in a subcategory of a larger herding group which contained other Collies and Shepherds (Figure 1). The Leonberger is an interesting breed to appear, as it is a large mastiff-type dog which was used as a cart-pulling dog in Germany.

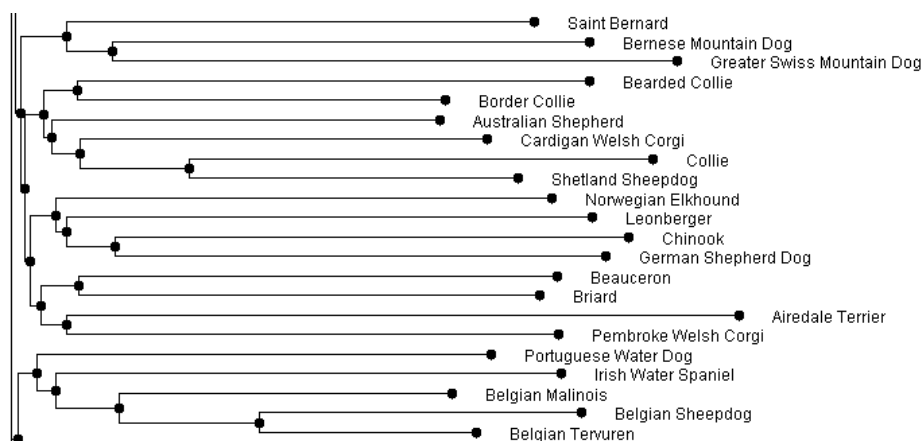


Figure 1: Chinook in the Neighbour-Joining tree.

The Chinook groups against all breeds:

This tree was similar to the Chinook by itself (Figure 1), though all the Chinook categories were clustered together and subcategories of Chinook (Pure) (P-P).

Different Chinook Samples against three breeds:

The tree for this case (Figure 2) clustered the crossbred Chinooks in the crossbreeding program into two groups with the three other pure breeds, Siberian Husky, Alaskan Malamute, and German Shepherd Dog between. The group on the left has a large amount of Chinooks with a German Shepherd Dog ancestor (G) with half the Pure Chinook (P-P) and Siberian Husky ancestry dogs (S). The group on the right has the Alaskan Malamutes ancestry dogs (M) with the rest of the Chinook samples, including the other half of the Siberian Husky ancestry dogs (S), and a couple German Shepherd Dog ancestry dogs (G). This shows quite clearly that the Chinooks are clustering with themselves, rather than as part of a different breed. This is important, as this is what other breeds do.

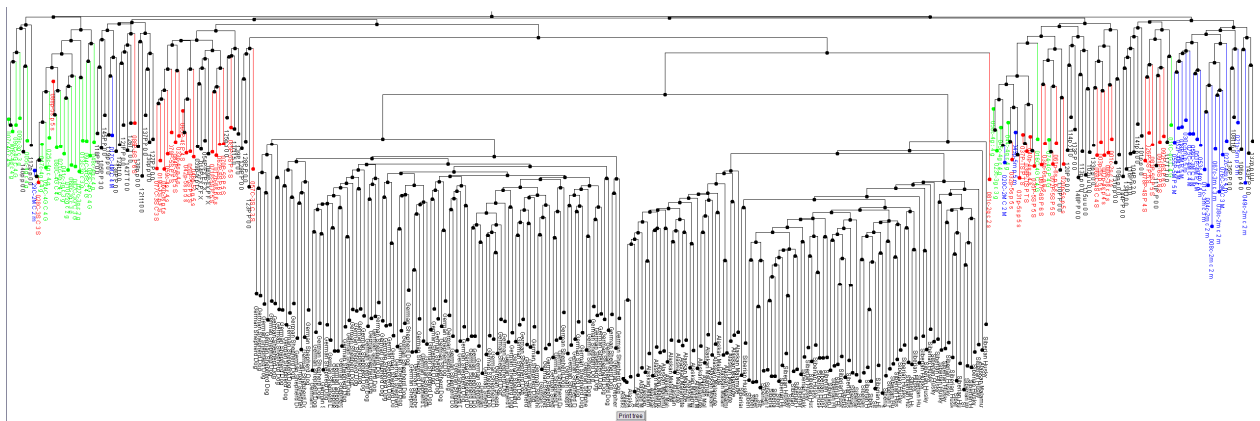


Figure 2: Dendrogram of Chinook samples with three breeds. Back-crosses of Siberian Husky Ancestry are in Red, Alaskan Malamute ancestry blue, and German Shepherd Dog ancestry in Green.

The Chinook samples without any other breeds:

When we removed the three other pure breeds from the Chinook samples, we got a more mixed tree (Figure 3), as the samples seem to be more dispersed. Some of the crossbred dogs from the breeding program are clustered together based on the crossbreed, but not all of them are, and the purebred dogs are scattered across all the dogs. For this reason, we would expect that this stratification is based mainly on the most recent purebred Chinook in the ancestry, rather than the crossbred. We think this is because the tree in Figure 2 was created with information about how close each sample was to the breeds in the initial cross of the back-crosses. This would cluster all Siberian Husky ancestry dogs together as the distance to Siberian Husky would be comparatively closer for all those back-crosses.

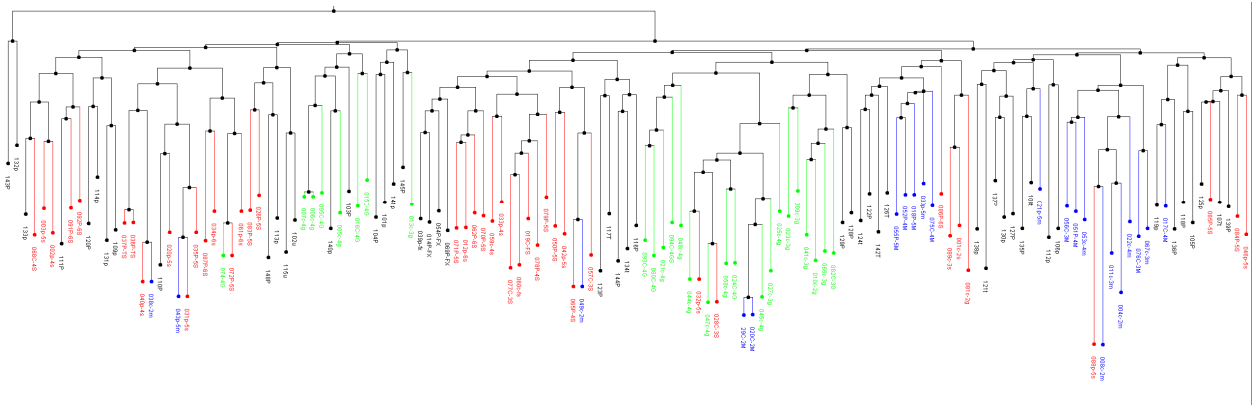


Figure 3: Dendrogram created for every Chinook sample.

Homozygosity in different Chinook groups

Since we have previously seen evidence of decreased average homozygosity in more cross-bred dogs, we were fairly sure what to expect in the various Chinook clusters.

In Table 10 we can see that the distribution of dogs in the crossbreeding program (CBP) tended to be less homozygous than the purebred Chinooks (UKC-P), and that the mean homozygosity of the dogs in the crossbred program which have become purebred Chinooks (C-P) is higher than in the other crossbred dogs with a more recent crossbred ancestor (C). It is interesting to note that all the UKC classified pures with a known cross-breed (T) have a higher average level of homozygosity than the rest of the pures (if it is a normal distribution). This suggests a more closed population, or possible inbreeding effect. It is possible that the distribution of the purebreds is binormal, caused by two different breeding populations in the data set, one with an average homozygosity between 0.65 and 0.675, and another population with an average homozygosity between 0.725 and 0.75. If this was the case, it might just suggest that the UKC classed purebreds with a crossbred ancestor (T) were crossed with only the more homozygous Chinook population.

	Crossbred			Pure		
	P	C		P	T	U
0.5	0	0		0	0	0
0.525	0	1		0	0	0
0.55	0	0		0	0	0
0.575	1	2		0	0	0
0.6	2	1		1	0	0
0.625	5	9		3	0	0
0.65	8	14		7	0	0
0.675	8	9		7	2	1
0.7	8	7		3	1	1
0.725	8	2		5	2	0
0.75	4	3		6	1	0
0.775	1	0		2	2	0

0.8	1	1		2	0	0
0.825	0	0		0	0	0
Sum	46	49		36	8	2

Table 10: Average homozygosity in dogs from different Chinook groups.

Table 11 shows the average homozygosity for dogs in the cross-breeding program at given generations back. The gradual trend towards higher homozygosity is to be expected in subsequent generations that have occurred since the back-cross event.

It may be an effect of the low resolution of samples, but it appears that the homozygosity splits into two groups at 6 generations back. A similar effect appears at 5 generations back, though less pronounced. This is similar to what we have seen in the possible pure set (P-P) above. We have drawn a possible divergence in the average homozygosity on a plot of the Table 11 data in Figure 4.

If this is a real, rather than a sampling effect, hypothetically what could be the reason? We know that a lower homozygosity occurs when you cross two individual breeds or populations. If the possible pure (P-P) was binormal, then these two different breeding populations could when crossed have a low homozygosity as seen. Rather than assuming two distinct Chinook populations, we could also explain this if the cross-breeding program only uses crosses with a partial subset of the Chinook population A, and then when the crosses reach 4 generations back, they are treated as purebred Chinooks and permitted to breed with the rest of the Chinook population B (causing the lower homozygosity), and either purebred dogs from the crossbreeding program, or dogs close to population A (causing the higher homozygosity).

	2	3	4	5	6
0.5	0	0	0	0	0
0.525	1	0	0	0	0
0.55	0	0	0	0	0
0.575	0	2	0	0	1
0.6	1	0	0	0	2
0.625	4	4	2	2	1
0.65	4	4	7	4	0
0.675	0	2	10	1	0
0.7	1	1	5	6	1
0.725	0	0	2	5	3
0.75	0	1	4	1	0
0.775	0	0	0	1	0
0.8	0	1	0	1	0

Table 11: Homozygosity in dogs (left) crossed at generations back (top).

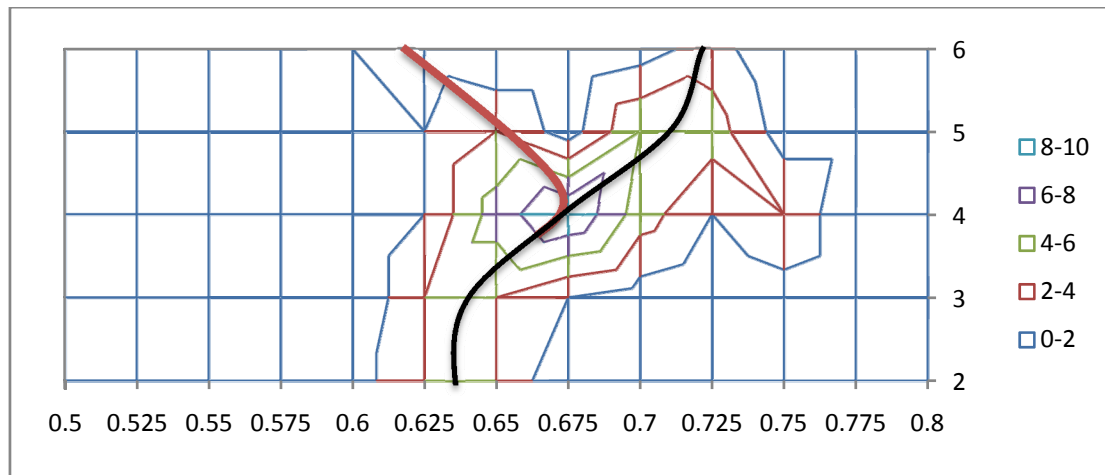


Figure 4: Possible divergence in average homozygosity in dogs crossed at generations back.

Homozygosity indicates many things in dog genetics. In what is called trait fixing, dogs can be bred so that you only breed dogs with a given preferred trait; coat texture, endurance or attitude for example. Over time these traits get fixed where these trait-specific genes are because we no longer have variants in the breeding population. Because of this, dogs will be homozygous at these locations. In the case of recessive diseases, it is more likely that a recessive disease is inherited if inbreeding between close relations occurs. Although the visible effect of both processes would be high homozygosity in a population, we are unable to determine which is the cause, as we are only concentrating our analysis around a small region which may be trait-linked, and not the entire chromosome. Given our knowledge of other breeds, however, we can say that based on our markers, the Chinook is by far not the most homozygous breed we have seen and would be mid-ranked in terms of average breed homozygosity at the Wisdom Panel SNP markers. This may be due in part to the variation allowed for various traits of the breed.

Analysis based on the Wisdom Panel

As previously stated, in the past year we have advanced our mixed-breed test into a swab test (Wisdom Panel Insights) and a more advanced blood test (Wisdom Panel Professional). As part of this improvement we added a Chinook signature to both tests after obtaining permission to do so. When this signature was tested, both products reported purebred Chinook for the input samples tested. This was to be expected as previous studies on the Chinook breed and cross-bred program dogs have suggested that these dogs form a distinct genetic cluster from other purebred breeds (Figure 5).

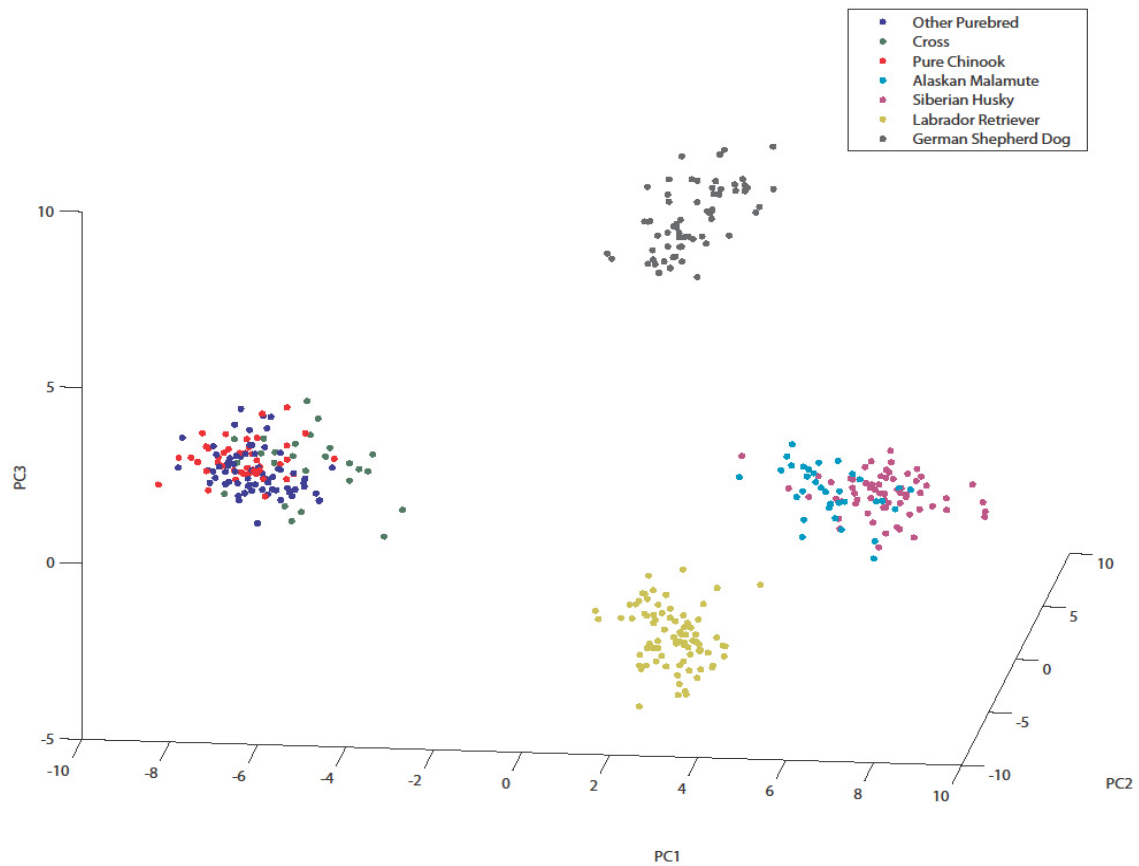


Figure 5: PCA of all Chinook samples vs other breeds.

Although this cluster is capable of identifying Chinook samples, it lacks finer detail in that it does not show any of the crosses we know have been used in the cross-breeding program (CBP) or in the UKC pures with crossbred ancestors (T).

To this end, we tried different combinations of Chinook samples to create clusters relevant to analyzing different artifacts in the dogs, including:

- No Chinook cluster: Designed to examine the Chinook makeup with the new algorithm and breeds.
- Possible Pure (P-P) Chinook Cluster: Designed to examine the back-crosses in the cross-breeding program (C).
- Pure dogs from the cross-breeding program (C-P): Designed to examine the cross-bred ancestors in the UKC classified purebreds (T).

We also used the Possible Pure (P-P) cluster to examine how the DIC statistic changes in different breed compositions.

Breeds detected in Wisdom Panel Professional without the Chinook cluster

Previous analysis using the Wisdom Panel algorithm on the Chinooks (without a Chinook Cluster signature) was completed on Wisdom Panel MX. Although Wisdom Panel Professional contains more breeds and has better detection of purebreds, the extra added breeds are fairly rare, so there is likely less difference to the results communicated previously. The purebred algorithm is also a factor which has a tendency to fit a purebred to dogs which appear homozygous enough to be pure; something which might be against the back-crosses and pure sets. We also added a Gray Wolf cluster to the algorithm for this analysis to see if it was a breed which was likely involved in the makeup of the Chinook.

The algorithm is a useful tool for examining the makeup of any breed potentially composited from multiple breeds as it is a mixed-bred test. Breeds which are not in the test will be generally replaced with a combination of breeds used in the creation of the breed, or breeds which are similar based upon statistical likelihood.

Table 12 shows the different proportions of breeds detected using the Wisdom Panel Professional with the Chinook cluster removed. The most dramatic observation is that the UKC pure (UKC-P) category has no Siberian Husky detected in the results, which is one of the most popular crosses in the crossbreeding program. One reason for this may be due to the purebred algorithm activating in 44% of the pure category (UKC-P), and 38% of the dogs from the cross-breeding program, and when active, replacing the entire dog with the closest purebred which explains the data.

Most of the other breeds were detected in similar measures when we compared between these two categories, which is to be expected when the vast quantity of genetic material in the dogs from the cross-breeding program (CBP) is Pure Chinook. It is interesting to see Norwegian Elkhound in this analysis, a dog which would have been used for sledding back in its native Norway since possibly even before the Vikings. Some of these dogs are likely to appear as false positives due to being more heterozygous, these are listed in red. Despite adding the Gray Wolf to the analysis, this only appeared in two results as a potential great-grandparent, but was not detected with a high enough confidence to be likely to be a true match.

	Popularity in Crossbred (CBP)	Popularity in Pures (UKC-P)
German Shepherd Dog	17.18%	19.59%
Alaskan Malamute	14.10%	18.56%
Collie	13.66%	15.46%
Norwegian Elkhound	6.17%	8.25%
Bouvier Des Flanders	6.61%	4.12%
Siberian Husky	5.73%	0.00%
Labrador Retriever	3.52%	3.09%
Plott	2.20%	4.12%
German Spitz	2.20%	2.06%
Russell Terrier	1.76%	2.06%
Icelandic Sheepdog	2.20%	1.03%
Anatolian Shepherd Dog	2.20%	1.03%
Australian Cattle Dog	1.32%	3.09%
Belgian Malinois	1.32%	2.06%

Table 12: Popularity of breed in various results for the top 14 most popular breeds.

Creating Alternate Wisdom Panel Clusters

To analyze the breed makeup further, we created another two clusters. The first was comprised of only UKC purebred Chinooks (P-P, U) without the AKC purebreds with known crossbred ancestors (T). This was designed to show if the test detected any amount of the admixed breeds in the initial cross of the back-crossed dogs in the cross-breeding program (C).

The second cluster was built from Chinooks from the cross-breeding program classed as pure with a crossbred ancestor at 5 or 6 generations back (C-P). This was designed to show any irregularities within the purebred group itself, specifically signs of German Shepherd Dog in the UKC pures with the known crossbred ancestors (T).

Table 13 shows the results for the first cluster. This shows that likelihood of detection of an admixed breed in the initial cross of the back-crosses decreases as the dog is further back-crossed. This is likely an effect of the dilution of chromatids remaining from the initial cross, as predicted previously in Table 2. We can also see that dogs with purebred Siberian husky as the initial cross (S) are detected more than four times as often as the back-crosses with the mixed Alaskan Malamute (M) and German Shepherd Dog (G) groups, though the proportion of Siberian Husky back-cross samples is only twice as many (Table 1). This is likely due to the fact that the initial cross is pure for the Siberian Husky group (S), yet more complex for the mixes (M, G) and hence more genetic material was segregating from that breed.

Generation	S	M	G	detected	supposed	accuracy at distance back
2	7	3	1	11	11	100%
3	5	1	0	6	15	40%
4	3	0	1	4	30	13%
5	1	0	0	1	21	5%
6	0	0	0	0	8	0%

Table 13: Detection of cross-bred ancestor at various ancestors back.

The second cluster managed to detect German Shepherd Dog in 4 out of 8 dogs (117T, 121T, 124T, and 134T), in the UKC pures containing a cross-bred ancestor (T) as well as detecting Alaskan Malamute in one of these dogs. It is interesting that it is more difficult to detect the Alaskan Malamute than the German Shepherd Dog in the 50/50 Alaskan Malamute / German Shepherd Dog cross in these dogs. This suggests that the genetic nature of the second Chinook cluster may have an Alaskan Malamute component which masks this breed from being detected reliably. This cluster also reported more breeds for samples from the cross-breeding program of 2 and 3 generations back, but reported crosses of 4, 5, and 6 generations back as pure Chinook.

These multiple clusters offer different utility for the Chinook club. The existing cluster should work well for detecting Chinooks, though the other two clusters may be useful for verifying cross-breeding practices or to see how mixed the analysis believes a dog to be; perhaps adding a 50/50 Chinook found at the pound access to the cross-breeding program, and then permitting it to the pure population when it has been sufficiently back-crossed.

It may be worth noting that during development testing and observation of mixed breed dogs in the pet population, the Chinook signature is rarely detected in mixed breed dogs, probably due to the rarity of the breed and uniqueness and clarity of the breed signature.

DIC in the different Chinook groups

The Deviation Information Criterion (DIC) is a statistic we use to choose between alternate models of different complexity in our algorithm. It is a number which tends to be low in less complex (e.g. Purebred) dogs and high in cases where the dog is very mixed or complicated in terms of genetic heritage.

We decided to analyze DIC in the different groups based on the Wisdom Panel Professional analysis with results generated with the first Chinook cluster comprised of only possible pures (P-P).

Table 14 shows this in more detail with the crossbred dogs with a cross-bred ancestor further back has a trend of lower DIC. This table also suggests that the UKC dogs with a known cross-bred ancestor (T) may be less complex than some of the purebred dogs moved from the cross-breeding program. It is surprising that the UKC pures of undisputed ancestry (U) have such a comparatively high DIC compared to the UKC pures which are possibly pure (P-P). Although we should not draw conclusions from a sample of two dogs, this may suggest that DIC is a statistic which defines how close a specific dog is to the cluster used, rather than how close it is to the actual breed.

	ALL		2	3	4	5	6	P-P	T	U
105	1		0	0	1	0	0	0	0	0
110	0		0	0	0	0	0	0	0	0
115	1		0	0	0	0	0	1	0	0
120	1		0	0	0	0	0	1	0	0
125	2		0	0	0	0	0	1	1	0
130	8		0	0	0	2	0	5	1	0
135	8		0	0	0	1	0	5	2	0
140	11		0	0	0	3	1	7	0	0
145	15		0	0	4	2	2	5	1	0
150	21		0	1	5	2	1	8	1	2
155	16		1	2	6	2	2	2	0	0
160	14		0	0	6	3	1	1	1	0
165	19		3	5	5	4	0	0	0	0
170	8		0	2	3	0	1	0	1	0
175	4		0	1	0	2	0	0	0	0
180	4		2	2	0	0	0	0	0	0
185	5		3	1	0	0	0	0	0	0
190	0		0	0	0	0	0	0	0	0
195	1		1	0	0	0	0	0	0	0
200	2		1	1	0	0	0	0	0	0

Table 14: DIC in dogs tested on a cluster made from Purebred dogs in the P and U categories.

Discussion:

This analysis has shown how the Chinook is a unique population of dogs whose individual components clusters together (in preference to clustering with other dog breeds) in all genetic analysis performed, even with dogs from the cross-breeding program. The most related dogs to the Chinook tend to be European-derived herding breeds and Nordic Spitz breeds. It was surprising to see a lack of Siberian Husky detected in the purebred dogs using the Wisdom panel, though it has to be said that Huskies and Malamutes are genetically much closer to each other than they are to any other breed, and these breeds could have confounded each others detection in some of the analysis.

The variety of the herding breeds seen throughout the analysis, such as the German Shepherd, Collie, Bouvier des Flanders, and the Belgian Sheepdog and Tervuren suggests that the Chinook may share common ancestors with one or more of these breeds. It is also interesting that two breeds that pull things, the sledding Norwegian Elkhound and the cart-pulling Leonberger have appeared in some analysis to be genetically similar to the Chinook, again potentially reflecting some common ancestral breed origins.

We have looked at various analyses which focus on highlighting the differences between the Purebred Chinooks and the cross-breeding population. We have seen how this difference is quite small, mainly because of the speed at which admixed genetic information can be lost during the backcrossing process. The most effective method of increasing diversity would be to increase the size of the breeding population, something which should keep the amount of potential haplotypes in the next generation as high as possible. There is a potential gain of diversity by using the crossbred dogs, but due to the rapidity at which genetic material is drained through backcrossing with a pure breed, this gain of diversity is likely to be mostly Chinook haplotypes, and has the effect of increasing the breeding population. If it was a goal of the cross-breeding program to increase the amount of traits from other breeds or crosses to the Chinook, we would suggest that these were introduced to the purebred population at an earlier generation before the majority of added genetic material was lost.

It is interesting to compare the ability of the Wisdom Panel Professional to detect a cross-bred ancestor at a given distance back. The accuracy of detection was very high in the dogs with a recent cross, but less in crosses further back (especially 4-6 generations). This is most likely an effect of the loss of genetic material. We should also point out that since we use 25/38 chromosomes, it is possible that there were chromosomes inherited from the cross-breed, but not on the chromosomes we focus on. It also shows how much more challenging mixed-bred ancestors are in the initial cross of back-crosses, especially when breeds in the mixes have genetic similarity to the breed we are distinguishing them from.

We have also seen some good consistency from the analysis which shows a high standard of record-keeping in the cross-breeding program. We have seen shared haplotypes between Crossbred Chinooks and the breeds in the initial crosses, we have seen the initial cross coming up consistently in analysis, and we have also seen evidence of the reported Alaskan Malamute / German Shepherd mix in the UKC purebreds with a known crossbred ancestor (T).

The cross-breeding program has been useful to deepen the understanding of homozygosity within dogs with real cases of managed outbreeding and backcrossing into the population. As part of our work, we understand that crossbred dogs are on average more heterozygous, and purebreds are generally on average more homozygous, and it has been interesting to see the rate at which this changes within the cross-breeding program, and how it appears to diverge in the purebred dogs in the cross-breeding program (C-P). We have also found it useful to watch how the DIC measure changes with background complexity.

Final note with regard to future developments: we are also planning some online diversity tools for a different breed club. We hope to make this so that breeders and owners can see the diversity of their dogs within a breed, and find out the potential diversity of pairings. At the moment, this is in a very early stage of development, but we feel the Chinook club may get some value out of this. Is there a contact at the breed club who would be willing to liaise and occasionally test this site during development and provide feedback and suggestions?