Phylogenetic Relationship of Dog Breeds

Zach Ryder

**Introduction**

Dogs are a unique species in that they have been subject to breeding for an extensive period of time and have resulted in a number of breeds that exceeds many other domestic animals. Where dogs’ domestic relationship with humans began as a bodyguard to nomadic groups of humans, their relationship has branched into a number of other niches. While some dogs are still used as bodyguards, they now fill roles like companion and coworker among others. As their role in human society has changed, humans have also changed dogs. Smaller, leaner builds to help hunt small animals, aggressive attitudes to aid in warding off predator, or diminutive sizes to fit into a purse are just a few examples of the way humans have changed dogs to fill roles. With all of this diversity, the question of how it arose becomes pressing. It is common knowledge that a common ancestor between dogs and wolves was the point at which humans began breeding dogs. Where and when specific dog breeds arose are harder questions to answer.

An ongoing project at the NIH, titled the Dog Genome Project has attempted to elucidate the relationship of over 150 dog breeds. They utilize SNP data in their analysis and report 23 clades of related breeds. Given the variability of SNPs in a population, this choice makes sense. That being said, analyzing other genomic information may result in a different topology. The difficulty of using another part of the genome in determining phylogenetic relationships is the similarity of those segments between different dog breeds. Although a chihuahua and a great dane may look vastly different, the are still similar on a genetic level as they are still the same species. This leads to vast sections of the genome being incredibly similar where, if different species were being compared, would result in points of difference to make comparisons. However, if a portion of the genome was able to be used to make valid comparisons, it would be useful in corroborating the proposed phylogeny that the Dog Genome Project has proposed.

**Methods**

To take a different approach in studying the phylogenetic relationship among dog breeds, portions of the mitochondrial genome of 106 dogs were obtained. The sequences were obtained from GenBank and were aligned using MAFFT. Some breeds had multiple samples, but only one sequence for each breed was used. 35 of the sequences were full mitochondrial genomes while the remaining were different portions of the mitochondrial genome. The wolf (*Canis lupus lupus*) was used as an outgroup to help root the tree due to its known relationship with dogs. This construction of the data set was finalized with 28 unique breeds including the wolf. Subsequent analysis was performed using RAxML-NG to determine the phylogenetic relationship of the samples using maximum likelihood and bootstrapping. Several different models were used to compare likelihoods (GTR+G, GTR+G+FC, and GTR+RO+F4) in an effort to find the model that produced the best likelihood. Analysis was performed by generating 20 maximum likelihood trees for each model and then running 1000 TBE bootstraps.

**Results**

Figure 1 displays the tree determined to be the most likely. It was found using GTR+R4+FO as a model. This model resulted in the highest log likelihood with a value of -26093.755780. GTR+FC and GTR+FO had log likelihood values of -26551.965905 and -26551.530267 respectively. The GTR+FC, GTR+FO, and GTR+FO+R4 runs converged after 750, 600, and 350 bootstraps respectively.

**Discussion**

As shown in Figure 1 and 2. The tree obtained from mitochondrial genomic data differs from the tree obtained by the Dog Genome Project which used SNP data. There are several potential reasons why differences would arise between the two trees. A possible reason for these differences is that the breeds have mitochondria that are too similar genetically to get accurate isolation. It could be that the breeds are so genetically similar that any given breed is almost similar as any other breed and that the topology obtained isn’t particularly meaningful. The prior research was done using SNP data which presumably has more variable sites to infer relationships from and may have a higher mutation rate than mitochondrial DNA. This would result in analysis from SNP data suggesting that the breeds are very different while the mitochondrial data would suggest otherwise. Another possible reason for the difference between the two trees is the sources of the data. As mentioned previously, the data for this study was obtained from sequences on GenBank. The sequences used to make the other tree were obtained by attending dog shows and traveling to other similar locations to find pure breeds of dogs that could be sampled. The difference between these two sources could matter in that it is harder to know whether the sequences from GenBank came from a pure breed. In dog shows, the owners often have to have a pedigree or similar document that shows their dog is the type they say it is. GenBank does not have this requirement and as such, may have data that does not come from a true pure breed. This problem leads to a final problem: cross-breeding. Because dog breeds are still within the same species, different breeds could have been bred together to make another breed. This would lead to a genetic web where rather than a traditional tree, the breeds would have a lot of connections that cross between clades or populations that make it harder to discern clear relationships. These three problems could all be limitations in this study and, if remedied, could lead to more accurate conclusions. If these differences are truly due to the source of the DNA then this indicates that the previously determined tree may not be as accurate as expected and additional analysis should be done to determine whether a better topology exists. Mitochondrial genomes may be better datasets than SNP data for determining phylogenetic relationships because they can be traced matrilineally whereas SNP data is not as easily traced back.

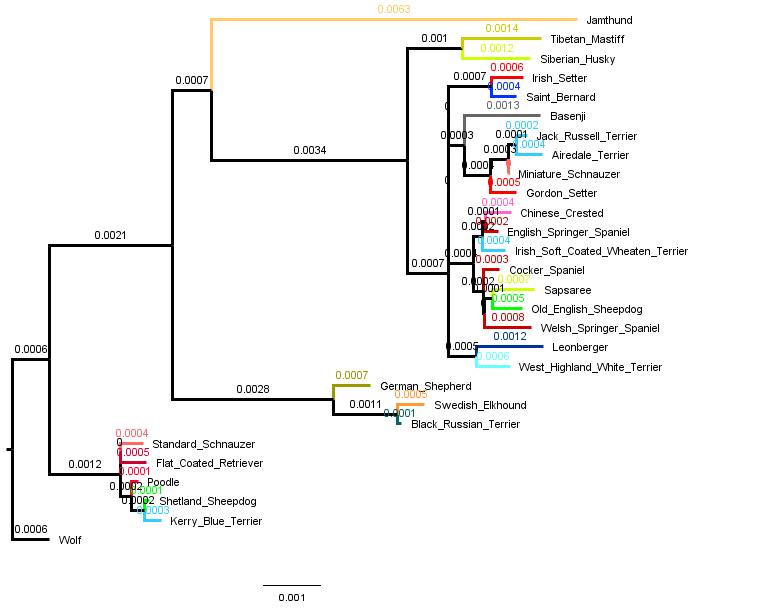
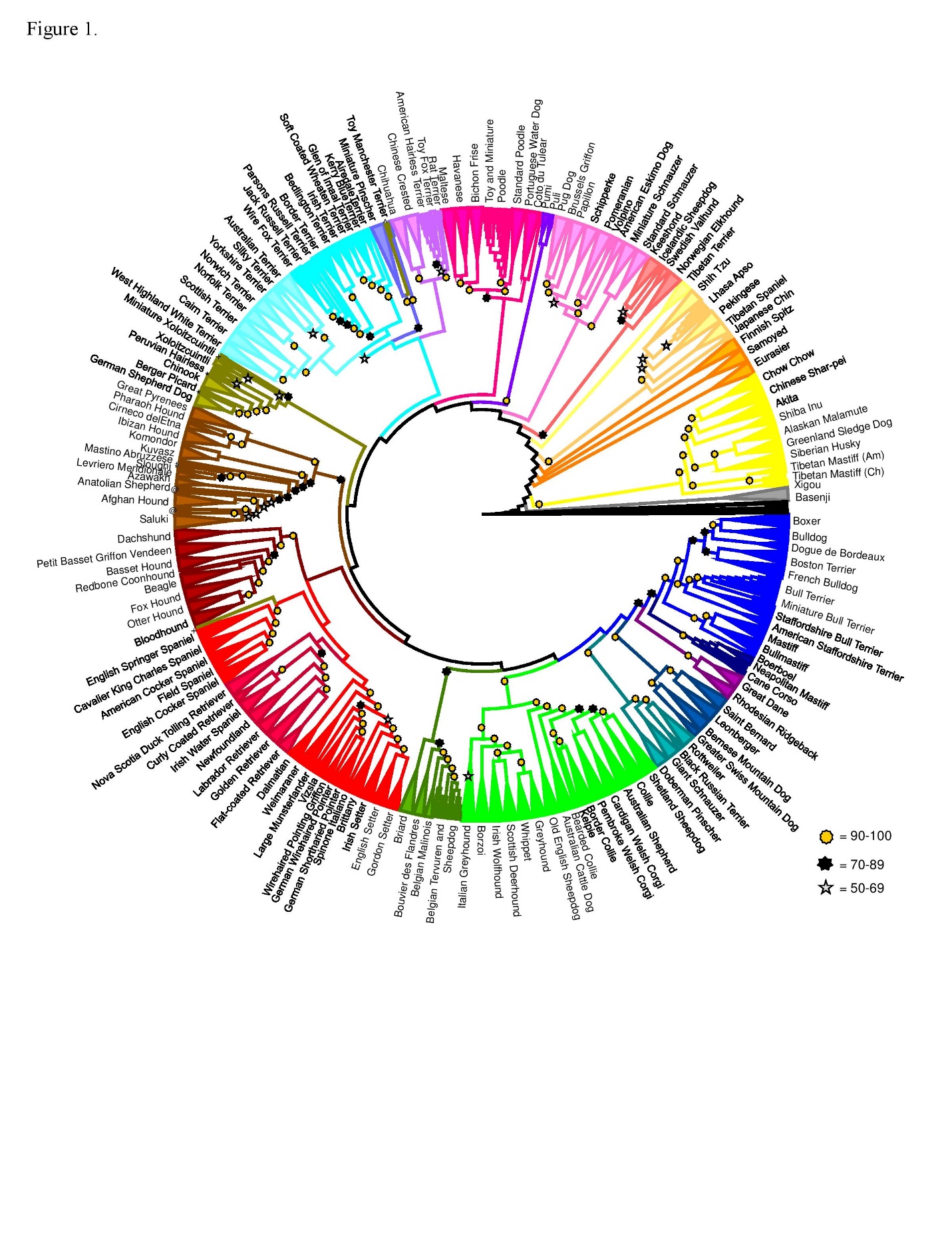


Figure 1. The phylogenetic tree obtained using RAxML-NG.

Figure 2. The tree determined by the Dog Genome Project.

**References**

Parker, Heidi G et al. “Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development.” *Cell reports* vol. 19,4 (2017): 697-708. doi:10.1016/j.celrep.2017.03.079