Phylogenetic Relationship of Dog Breeds

Zach Ryder

**\*\*\***I’m pretty sure I did something wrong with the analysis so I’m rerunning it but here are my results for now. Let me know whether I should be including more information. Definitely give as much criticism as you can, I’ll gladly take what I can get\*\*\*

**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 used. 35 of the sequences were full mitochondrial genomes while the remaining were different portions of the mitochondrial genome. The sequences were obtained from GenBank and were aligned using MAFFT. Some breeds had multiple samples and were signified with numbers (German Shepherd One, German Shepherd Two). Additionally, some breeds had full mitochondrial genomic data and partial sequences. These were denoted with the suffixes full and partial (Cocker Spaniel Full, Cocker Spaniel Partial). The wolf (*Canis lupus lupus*) was used as an outgroup to help root the tree due to its known relationship with dogs. 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 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 -27729.284303. GTR+G and GTR+G+FO had log likelihood values of -28167.807944 and -28167.794177 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. The mitochondrial tree has a few abnormalities that indicate either analysis needs to be performed again or that it should be performed for a longer amount of time. The primary abnormality is the wolf not being placed as the outgroup of the tree and instead being found in the middle of the topology. This placement is contrary to current historical opinion. Another abnormality is the lack of clades for the duplicate groups. It would be expected that the sequences which came from the same breed would end up forming a clade because it comes from the same animal. It should be noted that this could be due to the sample coming from a different part of the mitochondrial genome (the R-loop instead of the cytochrome-b region for example). A possible reason for these abnormalities is that the breeds are too similar genetically to get accurate isolation. It could be that the breeds are so genetically similar in the areas being analyzed that any given breed is just as similar as any other breed and that the topology obtained isn’t particularly meaningful. It should be noted that the branch lengths shown in Figure 1 are not scaled, the initial scaling resulted in flat tree with Pug Two having a massively long branch compared to all the other breeds. 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.

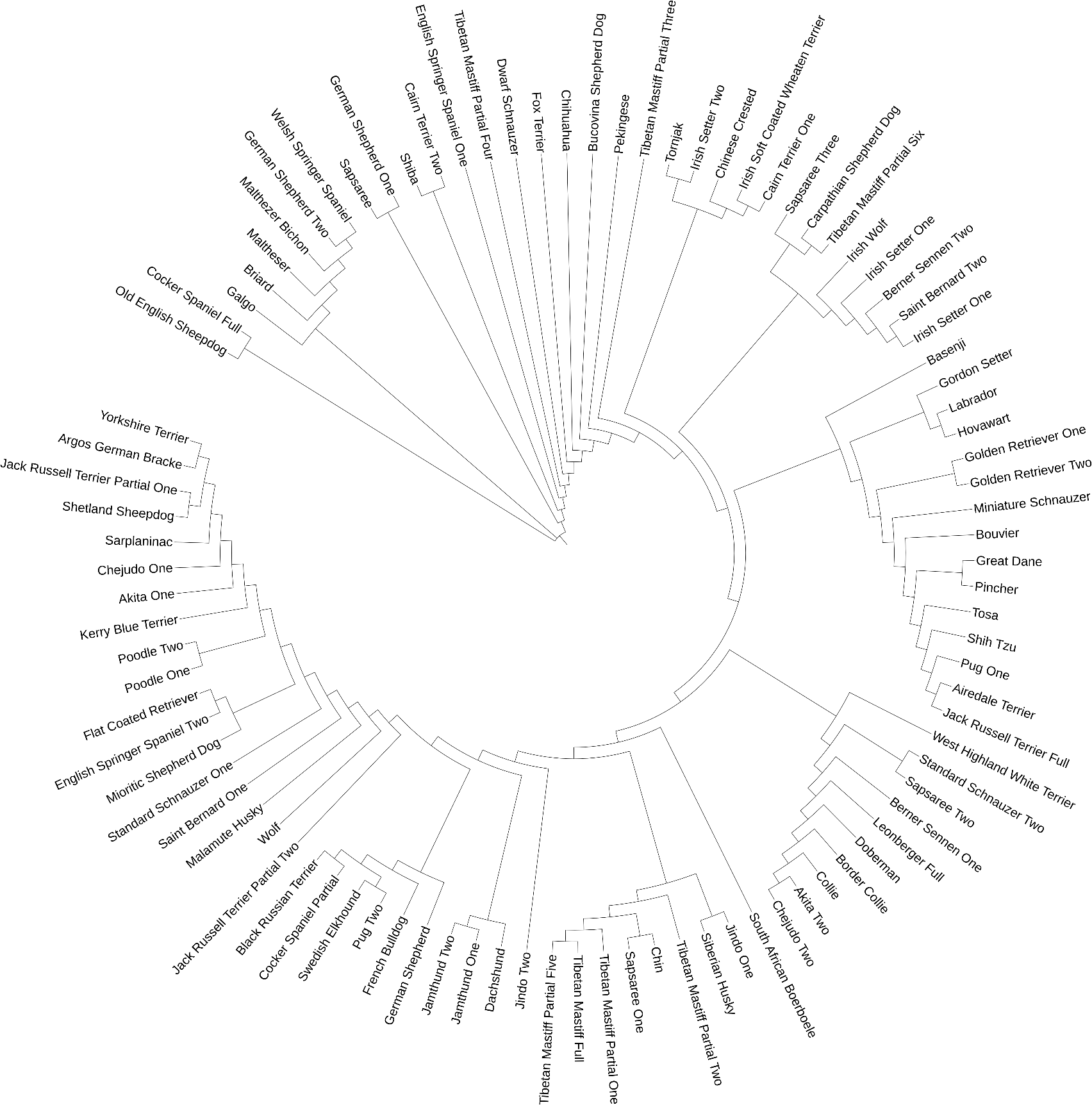
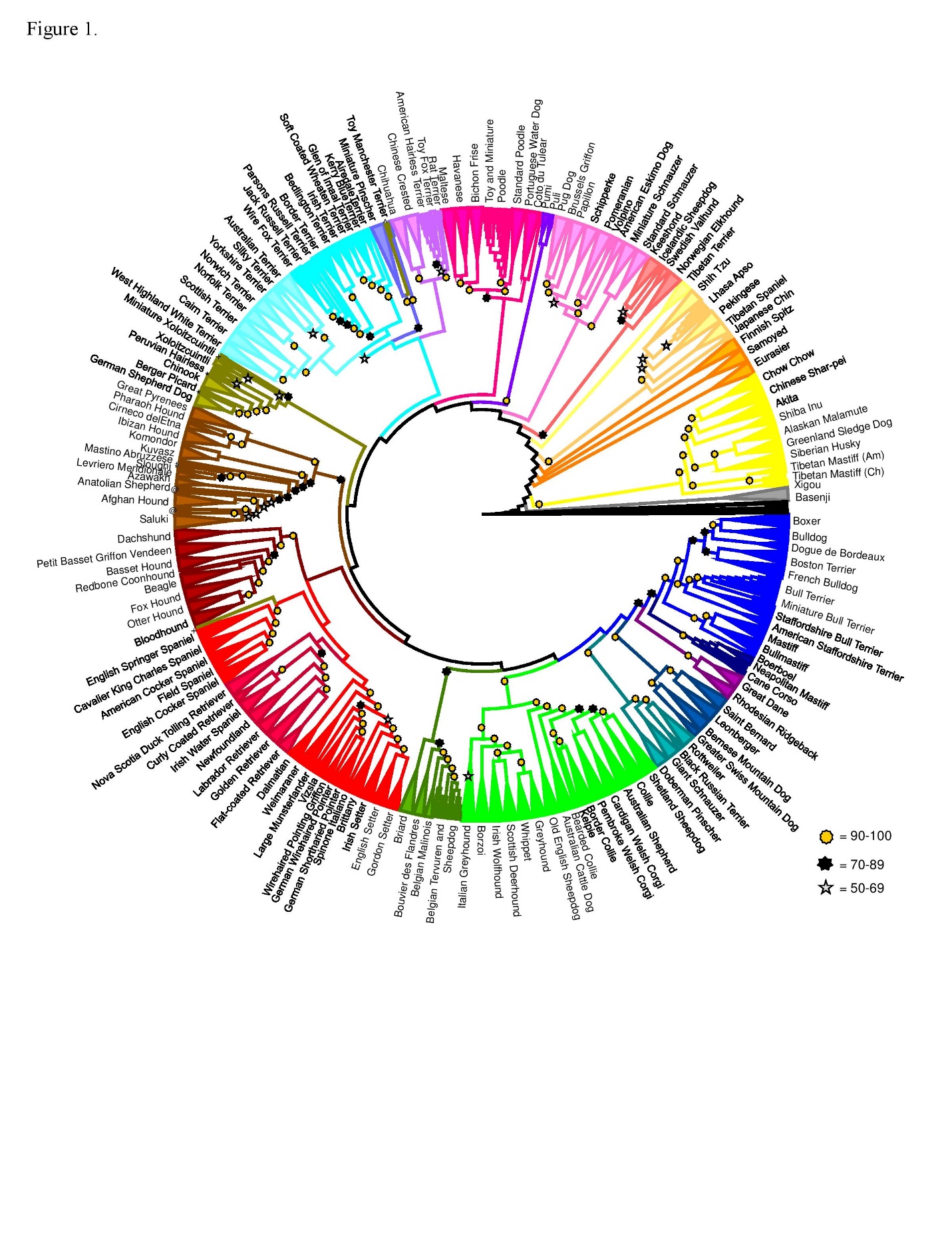


Figure 1. The phylogenetic tree obtained using RAxML-NG.

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