



## Phylogenetic Analysis of Maternal Lineages in Modern-Day Breeds of British *Canis lupus familiaris*

Tommy Rodriguez

Pangaea Biosciences, Department of Research & Development, Miami, FL USA

**\*Corresponding Author:** Tommy Rodriguez, Pangaea Biosciences, Department of Research & Development, Miami, FL USA

**Abstract:** Domesticated dogs are byproducts of controlled breeding practices that make it difficult to establish a historically accurate phylogeny of divergent events among any localized group. Despite the difficulties of arriving at an accurate inference, this study found distinct phylogenetic clusters within modern-day British dogs that largely corresponded to phenotype or function. Here, I examine maternal lineages among post-Victorian dog breeds by way of mitochondrial DNA (mtDNA) sequences. Two distinct sets of raw sequences were used in this investigation. My results will show that modern-day hounds, herding breeds, and spaniels fall nearest the midpoint of unrooted trees and contain the highest nucleotide substitution rates in the entire network, making these particular varieties strong candidates for the closest living relatives to the oldest lineages of European ancestry in Britain.

**Keywords:** domestic dog, phylogenetics, multiple sequence alignment, pairwise sequence alignment

### 1. INTRODUCTION

The origin of domesticated *Canis lupus familiaris* is not clear. Its closest living relative is the gray wolf (*Canis lupus*), and there is no evidence of any other canine contributing to its genetic lineage (Frantz et al., 2016, Thalmann et al., 2013 & Vilà et al., 1997). Studies propose a divergence time of the dog from the wolf ancestor at 27,000 YA (Freedman et al., 2014 & Skoglund et al., 2015), with the most recent estimate of domestication occurring between 20,000 and 40,000 YA (Botigué et al., 2017). The cohabitation of dogs and humans would have greatly improved the chances of survival for early human groups, and the domestication of dogs may have been one of the key forces that led to human success (Newby, 1997). The oldest dog breeds evolved or were bred to fill certain roles (Parker et al., 2017). A recent paper published in the journal *Science* said domestication likely happened from two separate wolf populations, one in Europe and the other in Asia (Coustal, 2017). British breeds are of a particular interest, due to their many well-established varieties.

A robust phylogeny of post-Victorian era British dog breeds has not yet been established. Other phylogenies have looked at the relationship between modern-day dog breeds on broad scales, including European varieties, but neglect to detail geographically isolated populations. This study focuses primarily on comparative techniques in phylogeny research to best infer the earliest members of, or most closely related modern-day breeds to the original varieties on the British Isles. Here, I examine maternal lineages among post-Victorian dog breeds by way of mitochondrial DNA (mtDNA) sequences. Two distinct sets of raw sequences were used in this investigation; one consists of complete genomes, whereas the other set is made up of partial sequences. My results will show that modern-day hounds and herding breeds fall closest to the midpoint of unrooted trees, making these particular varieties the closest living relatives to the oldest living lineages of European ancestry in Britain.

### 2. METHODS

#### 2.1. Sequence Selection

The NCBI nucleotide databank is the repository where each mtDNA sequence was acquired. BLAST similarity searches regularly facilitated identifying homologous sequence candidates among closely related groups. Each set of raw sequences are referenced in 4 studies: (1) Sequence Diversity of the

Canine Mitochondrial Genome (Shahid et al., 2004); (2) Mitochondrial genome DNA analysis of the domestic dog: identifying informative SNPs outside of the control region (Webb et al., 2009); (3) Identification of Single Nucleotide Polymorphisms within the mtDNA Genome of the Domestic Dog to Discriminate Individuals with Common HVI Haplotypes (Imes & Sacks, 2011); and (4) Forensic Informativity of ~ 3000 bp of Coding Sequence of Domestic Dog mtDNA (Angleby et al., 2014). From these collective findings, 2 distinct FASTA files containing a combination of 36 mtDNA sequences were compiled. Among the 36 mtDNA sequences, 15 sequences were available in complete format; while the remaining sequences were collected in partial format. See Table 1 for references, annotation numbers, and sequence descriptions.

## 2.2. Multiple Sequence Alignment toward Phylogenetic Reconstruction

UniPro UGENE v.1.26<sup>1</sup> was utilized in multiple sequence alignment (MSA) toward phylogenetic reconstruction, where Kalign was the primary source of algorithmic selection. An accurate and fast MSA algorithm, Kalign is a dependable algorithm for purposes of obtaining highly-robust base-pair alignments (Lassmann & Sonnhammer, 2005). Kalign is an extension of Wu-Manber approximate pattern-matching algorithm, based on Levenshtein distances. This strategy enables Kalign to estimate sequence distances faster and more accurately than other popular iterative methods. Lassmann and Sonnhammer (2005) show that Kalign is about 10 times faster than ClustalW and, depending on the alignment size, up to 50 times faster than other iterative methods; Kalign also delivers better overall resolution (Lassmann & Sonnhammer, 2005).

During this phase, Kalign for MSA gap penalty scores were modified slightly during successive intervals until a pair of optimal global alignments were achieved, including: (a) Complete mtDNA Sequences (16,784 bp); and (b) Partial mtDNA Sequences (16,742 bp). Next, I implemented PHYLIP neighbor-joining method coupled with distance matrix model F84 on both sets of base-pair alignments. This procedure would require additional bootstrapping compilers to help evaluate the strengths of the inner and outer nodes. Each tree-building exercise produced an individual phylogenetic tree for later analysis. The results of pairwise identity ratios are also outlined below.

## 2.3. About Algorithmic Selection for Tree-Building Execution

PHYLIP neighbor-joining is an accurate and statically consistent polynomial-time algorithm that does not assume that all lineages evolve at the same rate, and it constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join [where a set of  $n$  taxa requires  $n - 3$  iterations; each step is repeated by  $(n - 1) \times (n - 1)$ ] (Felsenstein, 1981 & Ragan, 1993). Suitable for inter-species comparison such as these, PHYLIP neighbor-joining is effective for producing highly probable diagrams amid scenarios involving low degrees of variance, regardless of alignment size. This method utilizes a set of default parameters for distance matrix model F84. Additional bootstrapping compilers were not required for this operation, and transition ratios are generated automatically under default settings. For reference purposes, the following formula demonstrates a standard neighbor-joining Q-matrix algorithm:

$$Q(i,j) = (n - 2) d(i,j) - \sum \{n, k = 1\} d(i,k) - \sum \{n, k = 1\} d(j,k) \quad (1)$$

Pair to node (distances):

$$(f,u) = \frac{1}{2} d(f,g) + \frac{1}{2}(n - 2) [ \sum \{n, k = 1\} d(f,k) - \sum \{n, k = 1\} d(g,k) ] \quad (2)$$

Taxa to node (distances):

$$d(u,k) = \frac{1}{2} [ d(f,k) + d(g,k) - d(f,g) ] \quad (3)$$

## 3. RESULTS AND DISCUSSION

### 3.1. Unrooted Trees

Domesticated dogs are byproducts of controlled breeding practices that make it difficult to establish a historically accurate geographical point of origin, even among the oldest lineages on a phylogenetic tree. Some statistical evidence suggests that Asian and Middle Eastern varieties diverged long before what has been called the ‘Victorian Explosion’ of dog breeds in Britain and the rest of Europe (Parker et al., 2017). One such study identified 9 dog breeds that could be represented on the outgroup of a broad-scale inter-species phylogeny (Pollinger et al., 2010). Pollinger et al. (2010) examined 48,000

single nucleotide polymorphisms that gave a genome-wide coverage of 912 dogs representing 85 breeds. Among the oldest lineages of European descent were those most closely related to herding breeds, mastiffs, and other hound varieties (Parker et al., 2017 & Pollinger et al., 2010).

Up until now, the quality of a detailed geographical analysis is sparse and unclear. Examining a localized dog population within continental Europe presents a challenge to a phylogenetic assessment, due to cross-breeding occurrences beyond regional boundaries. Such phylogenies are impractical because they do not resolve any evolutionary derived clades, as a result of artificial selection practices. What's more, it is unlikely that a single source in the form of a common ancestor could be identified among European breeds, beyond *Canis lupus*. When a single ancestral root cannot be inferred nor assumed, a traditional cladogram does not provide the most ideal resolution. Instead, we could illustrate the relatedness of the leaf nodes without making assumptions about deep ancestry.

Britain's unique geographical features make it a particularly interesting candidate for this type of phylogenetic analysis. The coastline and landscape of what would become modern-day Britain began to emerge at the end of the last Ice Age around 10,000 YA (Lane, 2011). What had been a cold, dry tundra on the north-western edge of Europe grew warmer and wetter as the ice caps melted (Lane, 2011). It wasn't until 6100 BC that Britain broke free of mainland Europe, during the Mesolithic period (Lane, 2011). The earliest evidence of human presence in Britain is dated to 10,500 BC (Bradley, 2007 & Hammond, 2008). By around 4000 BC, the island(s) became populated by people with a Neolithic culture that had already been engaging in dog domestication (Darvill, 2010). The oldest British dog remains were found at Star Carr Yorkshire and are dated at 7538 BC (Day, 1996).

We might then imagine a scenario where the first domesticated dogs of Britain experienced a form of reproductive isolation from a secondary population on the mainland; this was initial and not permanent. Over time, isolation can exert unique evolutionary forces that result in the development of a distinct genetic reservoir (Nqayi, 2014). Geographical barriers would not prevent the later influxes of foreign cultures to the British Isles, including the Romans, Anglo-Saxons, and Vikings, just to name a few. Late arrivals introduced foreign dog breeds to the native population; and this too has been well-documented by historical archives (Morgan, 2010, Salway, 2001 & Yalden, 2010). Alas, we now have dog varieties that are cross-breeds between one and many distinct others, and not direct decedents of one primary lineage of the original inhabiting population.

This brings me back to my position regarding the best approach for illustrating a phylogeny. As it happens to be in this particular case, unrooted trees best fit the profile for a phylogenetic assessment. This method can be generated from rooted trees by simply omitting the root (Maher, 2002). An unrooted tree should give no information about the order of speciation events (Garamszegi et al., 2014). Because the scope of this investigation does not necessarily rely on inferences regarding evolutionary history through divergent events, I look instead for patterns of relatedness among clustered networks in order to identify the sequence(s) with the highest nucleotide substitution rates. In terms of unrooted trees, this would be represented by the leaf node(s) most closely located to a midpoint on the diagram.

The unrooted tree(s) based on PHYLIP neighbor-joining polynomial values (Fig. 1 & Fig.2) illustrates a clear divide between different groups that reside on each end of the midpoint. My results found a distinct phylogenetic clustering within modern-day British dogs that largely correspond to phenotype or function; with two exceptions. That is to say, most terriers are grouped with other terriers, hounds with hounds, spaniels with spaniels, and herding breeds reside together on both sides of the midpoint. Collectively, these phylogenies are divided into two overlapping networks of breed types: (1) terriers, mastiffs, and setters; and (2) hounds, herding breeds, and spaniels. As Figure 1 & Figure 2 show, 17 out of 19 clades could be correctly assigned to their breed based on their genotype alone. The node(s) most closely located to the midpoint on Figure 1 & Figure 2 belong to the group of hounds; including the beagle, greyhound, whippet, and the Scottish deerhound, respectively. It may be worth noting that early historical accounts describe native British dogs that are hound-like in appearance (Short, 2009 & Hornblower et al., 2014). For instance, small hounds are mentioned in the Forest Laws of Canute (Short, 2009 & Hannas, 1978). If genuine, these laws would confirm that beagle-type dogs were present in England prior to 1016, but it is likely the laws were written in the Middle Ages to give a sense of antiquity and tradition to Forest Law (Short, 2009 & Hannas, 1978).

Spaniels reside nearest the midpoint in proximity to hounds, but on the adjacent end of the diagram (Fig. 2). Theories on the origin of British spaniels span further back, as it is believed that Welsh



### 3.2. Pairwise Identity Ratios

Homologous sequences have identities, and degrees of conservation and similarities are quantitative. Thus, by combining the alignments and comparing each sequence against *Canis lupis* mitochondrion [complete genome], we have a quantitative reference in which to better understand the variation behind these unrooted trees. With respect to similarity ratios, my results (Tab. 1) demonstrate a present correlation among the overlapping networks represented on the diagram(s) above; where the beagle (99%), Old English sheepdog (99%), Shetland sheepdog (99%), and Welsh spaniel (99%) contain the highest nucleotide substitution rates in the entire alignment.

**Table1.** Results of pairwise sequence alignment

Annotation Number [breed]	Complete Sequence	Partial Sequence
KF857179.1 [gray wolf]	-	-
AY729880.1 [beagle]	0.99	-
EU408265.1 [Welsh Corgi Pembroke]	0.98	-
AY656742.1 [Old English sheepdog]	0.99	-
JF342896.1 [greyhound]	-	0.98
DQ480500.1 [Shetland sheepdog]	0.99	-
JF342891.1 [whippet]	-	0.98
JF342885.1 [Scottish deerhound]	-	0.98
AY656747.1 [Welsh Springer spaniel]	0.99	-
AY656744.1 [English Springer spaniel]	0.99	-
EU408274.1 [English mastiff]	0.98	-
FJ817363.1 [golden retriever]	0.98	-
KU291081.1 [Labrador retriever]	0.98	-
JF342887.1 [English cocker spaniel]	-	0.97
JF342867.1 [English setter]	-	0.98
EU408254.1 [basset hound]	0.98	-
JF342861.1 [Harrier]	-	0.97
JF342856.1 [Welsh terrier]	-	0.97
JF342888.1 [West Highland white terrier]	-	0.97
JF342843.1 [Scottish terrier]	-	0.98
JF342875.1 [Norfolk terrier]	-	0.98
JF342833.1 [Manchester Terrier]	-	0.98
JF342899.1 [border collie]	0.98	-
AY656751.1 [Gordon setter]	0.98	-
AY656741.1 [Irish setter]	0.98	-
EU408294.1 [pug]	0.97	-
EU408263.1 [Cavalier King Charles spaniel]	0.98	-
KU290898.1 [Staffordshire bull terrier]	0.98	-
JF342901.1 [bull terrier]	-	0.98
JF342832.1 [bullmastiff]	-	0.98

### 4. CONCLUSION

Natural selection and selective breeding reinforce certain characteristics in dog populations, giving rise to dog types and dog breeds. In Britain alone, a recurring theme of outside cultural influxes may have significantly altered the evolutionary trajectory of the native breed population at any given period. Despite the difficulties of arriving at an accurate inference, this study found distinct phylogenetic clusters within modern-day British dogs that largely corresponded to phenotype or function. The unrooted tree(s) based on PHYLIP neighbor-joining polynomial values illustrates a divide between two overlapping networks. Falling nearest the midpoint, hounds, herding breeds, and spaniels represent a strong possible candidate for oldest living lineages in Britain; with modern-day beagles, sheepdogs, and Welsh spaniels having the highest degree of pairwise similarity ratios compared against *Canis lupis*.

### NOTES

<sup>1</sup> UGENE was used in comparative sequence analysis. The DNA sequences noted above are in FASTA format. They were obtained from the NCBI database archives.

REFERENCES

- Angleby, H., Oskarsson, M., Pang, J., Zhang, Y. P., Leitner, T., Braham, C., & Savolainen, P. (2014). Forensic Informativity of ~ 3000 bp of Coding Sequence of Domestic Dog mtDNA. *Journal of forensic sciences*, 59(4), 898-908.
- Botigué, L. R., Song, S., Scheu, A., Gopalan, S., Pendleton, A. L., Oetjens, M., ... & Bobo, D. (2017). Ancient European dog genomes reveal continuity since the early Neolithic. *Nature Communications*, 8.
- Bradley, R. (2007). The prehistory of Britain and Ireland. *Cambridge University Press*.
- Coustal, L. (2017). Study throws dog domestication theories to the wolves. *Phys.org*
- Darvill, T. (2010). Prehistoric Britain. *Routledge*.
- Day, S. P. (1996). Dogs, deer and diet at Star Carr: a reconsideration of C-isotope evidence from early Mesolithic dog remains from the Vale of Pickering, Yorkshire, England. *Journal of Archaeological Science*, 23(5), 783-787.
- Frantz, L. A., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., ... & Tresset, A. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290), 1228-1231.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6), 368-376.
- Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., ... & Beale, H. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS genetics*, 10(1), e1004016.
- Garamszegi, L. Z., & Gonzalez-Voyer, A. (2014). Working with the tree of life in comparative studies: How to build and tailor phylogenies to interspecific datasets. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (pp. 19-48). Springer Berlin Heidelberg.
- Hammond, N. (2008). "Flint hints at existence of Palaeolithic man in Ireland". *The Times*.
- Hannas, C. A. (1978). An analysis of hunting and sporting scenes portrayed in the decoration of glass, pottery, and porcelain wares in the Brunnier Collection.
- Hornblower, S., Spawforth, A., & Eidinow, E. (Eds.). (2014). The Oxford companion to classical civilization. *Oxford Companions*.
- Imes, D.L. and Sacks, B.N. (2011). Identification of Single Nucleotide Polymorphisms within the mtDNA Genome of the Domestic Dog to Discriminate Individuals with Common HVI Haplotypes. *University of California*. Retrieved from NCBI Nucleotide Database.
- Ireland, S. (2008). "Chapter 15: Government, Commerce and Society". Roman Britain: A Sourcebook. Routledge Sourcebooks for the Ancient World (3rd ed.). *Taylor & Francis*. p. 216. ISBN 9780415471770. OCLC 223811588.
- Lane, M. (2011). The moment Britain became an island. *BBC History*.
- Lassmann, T., & Sonnhammer, E. L. (2005). Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC bioinformatics*, 6(1), 1.
- Maher, B. A. (2002). Uprooting the tree of life: a proposed theory has researchers debating life's origins--again. *The Scientist*, 16(18), 26-28.
- Morgan, K. O. (Ed.). (2010). The Oxford History of Britain. OUP Oxford.
- Nqayi, Z. (2014). The commemoration of International Day for Biological Diversity. *Department of Environmental Affairs*, Republic of South Africa.
- Newby, J. (1997). The pact for survival: humans and their animal companions. *ABC Books for the Australian Broadcasting Corporation*.
- Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G., & Ostrander, E. A. (2017). Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*, 19(4), 697-708.
- Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., ... & Bryc, K. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, 464(7290), 898.
- Ragan, M. A. (1992). Phylogenetic inference based on matrix representation of trees. *Molecular phylogenetics and evolution*, 1(1), 53-58.
- Renna, C. (2008). Herding Dogs: Selection and Training the Working Farm Dog. *Kennel Club Books (KCB)*. ISBN 978-1-59378-737-0.
- Salway, P. (2001). A history of Roman Britain. *Oxford Paperbacks*.

- Shahid, S.A., Xiao, Y., Khan, S., Feng, D., Johnson, G.S. and Ha, J. (2004). Sequence Diversity of the Canine Mitochondrial Genome. *University of Missouri*. Retrieved from NCBI Nucleotide Database.
- Short, R. (2009). King Canute and the wisdom of forest conservation. *Nature*, 462(7273), 567-567.
- 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(11), 1515-1519.
- Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L.,...& Napierala, H. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, 342(6160), 871-874.
- Vilà, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L.,...& Wayne, R. K. (1997). Multiple and ancient origins of the domestic dog. *Science*, 276(5319), 1687-1689.
- Webb, K. M., & Allard, M. W. (2009). Mitochondrial genome DNA analysis of the domestic dog: identifying informative SNPs outside of the control region. *Journal of forensic sciences*, 54(2), 275-288.
- Yalden, D. (2010). The history of British mammals. *A&C Black*.

**Citation:** T. Rodriguez, "Phylogenetic Analysis of Maternal Lineages in Modern-Day Breeds of British *Canis lupus familiaris*", *International Journal of Research Studies in Biosciences (IJRSB)*, vol. 5, no. 9, pp. 41-47, 2017. <http://dx.doi.org/10.20431/2349-0365.0509008>

**Copyright:** © 2017 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.