Molecular Therapy
Methods & Clinical Development
Original Article

Mechanism of Deletion Removing All Dystrophin Exons in a Canine Model for DMD Implicates Concerted Evolution of X Chromosome Pseudogenes

D. Jake VanBelzen,[1] Alock S. Malik,[1] Paula S. Henthorn,[2] Joe N. Kornegay,[3] and Hansell H. Stedman[1,4]

[1]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [2]Section of Medical Genetics, University of Pennsylvania School of Veterinary Medicine, Philadelphia, PA 19104, USA; [3]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA; [4]Corporal Michael Crescenz Veterans Administration Medical Center, Philadelphia, PA 19104, USA

Duchenne muscular dystrophy (DMD) is a lethal, X-linked, muscle-wasting disorder caused by mutations in the large, 2.4-Mb dystrophin gene. The majority of DMD-causing mutations are sporadic, multi-exon, frameshifting deletions, with the potential for variable immunological tolerance to the dystrophin protein from patient to patient. While systemic gene therapy holds promise in the treatment of DMD, immune responses to vectors and transgenes must first be rigorously evaluated in informative preclinical models to ensure patient safety. A widely used canine model for DMD, golden retriever muscular dystrophy, expresses detectable amounts of near full-length dystrophin due to alternative splicing around an intronic point mutation, thereby confounding the interpretation of immune responses to dystrophin-derived gene therapies. Here we characterize a naturally occurring deletion in a dystrophin-null canine, the German shorthaired pointer. The deletion spans 5.6 Mb of the X chromosome and encompasses all coding exons of the *DMD* and *TMEM47* genes. The sequences surrounding the deletion breakpoints are virtually identical, suggesting that the deletion occurred through a homologous recombination event. Interestingly, the deletion breakpoints are within loci that are syntenically conserved among mammals, yet the high homology among this subset of ferritin-like loci is unique to the canine genome, suggesting lineage-specific concerted evolution of these atypical sequence elements.

INTRODUCTION

Recent progress in vector-mediated gene therapy shows promise in the treatment of Duchenne muscular dystrophy (DMD).[1] However, as in the case of many genetic diseases, a protein is mutated or altogether absent, preventing the establishment of immunological tolerance to its wild-type form. Thus, gene therapies that deliver a transgene modeled after a wild-type protein may contain epitopes to which the patient's immune system lacks central tolerance, and, therefore, they risk inciting a deleterious host immune response.[2,3]

In addition, the sporadic and highly varied dystrophin mutations within the DMD patient population make evaluation of immune responses following treatment exceptionally challenging, as each patient's immune system may react differently to the peptide product of a recombinant transgene. Therefore, an animal model void of immunological tolerance to all dystrophin epitopes should provide the most sensitive prediction of immune responses to gene therapies.

Preclinical development of gene therapies for DMD has centered on the use of two naturally occurring animal models, the mdx mouse and the golden retriever muscular dystrophy (GRMD) dog, which are caused, respectively, by a nonsense mutation within exon 23[4] and a point mutation within the splice acceptor site of intron 6.[5] These mutations represent only a small portion of those seen in the DMD patient population; therefore, they cannot accurately predict the potential human immune responses to DMD gene therapies, and there are currently no primate models for DMD. Furthermore, naturally occurring exon skipping and stop codon readthrough can result in leaky dystrophin expression, as evidenced by revertant fibers,[6–8] and they may allow for the establishment of immunological tolerance to dystrophin during development in these animal models. Alternatively, dystrophin expression in revertant fibers could result in a primed immune response to dystrophin peptides, as was shown in humans.[9] Both of these outcomes convolute the interpretation of immune responses to newly produced proteins acting as neoantigens.

The German shorthaired pointer-muscular dystrophy (GSHPMD) is a recently described, naturally occurring dog model of DMD.[10] Western blot and fluorescence in situ hybridization (FISH) analyses suggest that the deletion in this model may encompass the entirety of the dystrophin gene.[11] Here we used a PCR approach to precisely define the deletion endpoints and to confirm the complete absence of the *DMD* gene by sequencing across the deletion. We found that the deletion spans 5.6 Mb and is remarkably similar to that of patient

Received 27 October 2016; accepted 7 December 2016;
http://dx.doi.org/10.1016/j.omtm.2016.12.001.

Correspondence: Hansell H. Stedman, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, 422 Curie Blvd., 709A Stellar Chance, Philadelphia, PA 19104, USA.
E-mail: hstedman@mail.med.upenn.edu

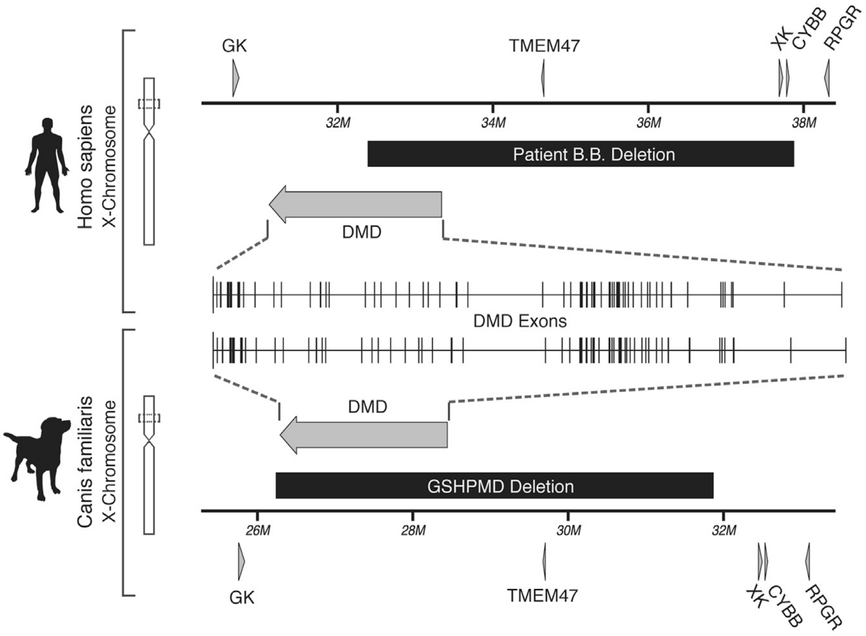


Figure 1. Comparison of Human and Dog *DMD* and Neighboring Genes

Orthologous regions of the human and dog X chromosome surrounding the *DMD* gene are compared. Arrows indicate genes, with the *DMD* gene expanded inward for comparison of exons between the two species. The X chromosome deletion in the historic patient, B.B., is depicted in a black box, as is the GSHPMD deletion. Arrowheads indicate the direction of transcription. GK, glycerol kinase; DMD, dystrophin; TMEM47, transmembrane protein 47; XK, X-linked Kx blood group; CYBB, cytochrome b-245 beta polypeptide; RPGR, retinitis pigmentosa GTPase regulator; M, million.

B.B., a boy whose deletion dramatically accelerated the characterization of the *DMD* locus (Figure 1).[12–16] Interestingly, the GSHPMD deletion breakpoints are within highly homologous DNA loci that are conserved on the mammalian X chromosome. The GSHPMD model, lacking the entirety of the dystrophin gene and, therefore, void of any possible level of immunological tolerance or sensitivity to wild-type dystrophin epitopes, could provide a much-needed platform for the prediction of immune responses to gene therapies for DMD.

RESULTS

Despite 100 million years of evolution,[17] the regions of the human and canine X chromosomes that encompass *DMD*, including exon-intron spacing and size, are conserved (Figure 1), perhaps indicative of the vital role its encoded peptide, dystrophin, plays in muscle biology. For clarity in the presentation of our results, we reference the default orientation and numbering of the X chromosome delegated by the NCBI, which assigns *DMD* to the antisense strand in the dog genome assembly (Figure 1).

Genetic Mapping of the GSHPMD Deletion Breakpoints

To map the deletion in the GSHPMD model, we designed a PCR-based strategy to locate the breakpoints on the X chromosome. Using previously reported FISH data from a GSHPMD carrier female,[11] we estimated the location of the deletion and designed primer pairs that broadly spanned this region of the X chromosome (Figure 2). Gel electrophoresis was used to compare PCR results from wild-type (WT) dog or GSHPMD DNA to broadly map the deleted region of the X chromosome. Following a similar approach, additional primer pairs were designed to finely map the telomeric (TBP) and centromeric (CBP) breakpoints. In this way, we employed 74 unique primer pairs and mapped the TBP to be within base pairs 26,237,921–26,239,551 and the CBP to be within base pairs 31,867,082–31,871,007 of the dog X chromosome, proving that the deletion spans 5.6 Mb and encompasses the entirety of the *DMD* and *TMEM47* genes.

Further attempts to more finely map the deletion breakpoints repeatedly failed (data not shown), perhaps due to a 711-bp genome assembly gap at base pairs 26,238,732–26,239,442 within the TBP region (Figure S1A). To determine the DNA sequence of the assembly gap, we used primer pair 74 to PCR-amplify the respective genomic region from a BAC clone from the library used in the dog genome assembly project (Figure S1B).[18] Amplification of this region was difficult and dependent on the addition of betaine, a PCR enhancer, which may explain why the region was not sequenced in the dog genome assembly. Following amplification, we gel-purified and sequenced the 1.5-kb PCR product and assembled the individual Sanger reads into a single DNA contig (GenBank: KR907258; Figure S1C). A Pustell DNA matrix was used to compare the assembled contig to the region of the X chromosome that harbors the assembly gap (Figures S1D and S6). This comparison revealed >97% homology between our sequenced PCR product and the chromosomal sequence flanking the genome assembly gap, suggesting that the internal 650-bp region of the PCR product sequence is representative of the previously unknown assembly gap sequence. We therefore replaced base pairs 26,238,718–26,239,484 of the dog X chromosome with base pairs 338–1,045 of our BAC-derived, PCR product, and we used the resulting sequence in our subsequent analyses.

The GSHPMD Deletion Spans 5.6 Mb and Is Contiguous

To confirm the absence of the entire 5.6-Mb DNA fragment from the X chromosome, we designed a primer pair that flanks the predicted deletion. The primer pair sequences encompass 5.6 Mb, which is far outside the limits of PCR. However, unique to PCRs containing DNA from affected GSHPMD males and carrier females, a 2-kb amplicon was generated, suggesting that the entire 5.6-Mb region is deleted from the X chromosome in the GSHPMD model (Figure 3A).

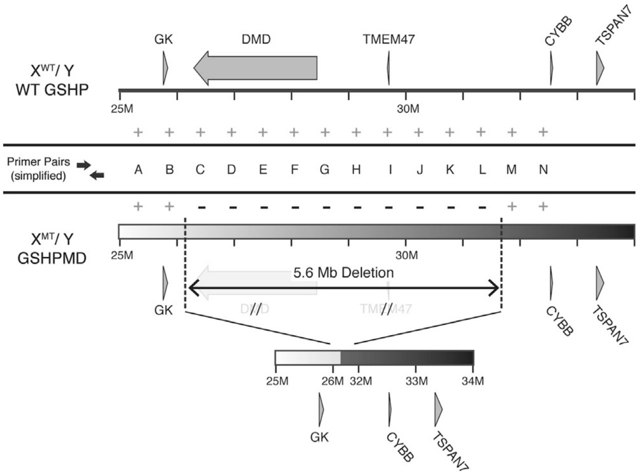


Figure 2. Deletion Map of the GSHPMD Model by PCR

The X chromosome of male wild-type GSHP and affected GSHPMD dogs are compared. The locations of primer pairs are labeled alphabetically. The results of each PCR experiments are shown, with plus indicating successful amplification and minus indicating failed amplification. The mapped deletion in the GSHPMD model is depicted inferiorly. X^{WT}/Y, wild-type male; X^{MT}/Y, mutant male; PCR, polymerase chain reaction; TSPAN7, tetraspan 7; Mb, million base pairs; M, million.

To confirm its identity, we gel-purified and sequenced the GSHPMD-specific amplicon, and subsequently we assembled the individual Sanger reads into a 1.7-kb contig (GenBank: KR907259). A Pustell DNA matrix comparing the deletion-spanning contig and base pairs 26–33 Mb of the dog X chromosome revealed >90% sequence homology between the 5′ portion of the contig and the TBP region, with an eventual, subtle shift in homology to favor the CBP region (Figure 3B), consistent with the sequenced amplicon-spanning deletion. A short break in homology near the TBP was present in our original Pustell matrix, but it was found to result from three insertions or deletions (indels) not accommodated by the matrix (Figure S2A). In addition, a break in homology attributable to a TAAA tandem repeat was present between 400 and 500 bp of the Pustell matrix. The abundance of this repeat became apparent when specificity parameters of the Pustell matrix were reduced (Figure S2B).

To our surprise, our Pustell matrix (Figure 3B) also revealed that the TBP and CBP regions were highly homologous to each other, raising questions as to the mechanism of the deletion. Importantly, both the 5′ and 3′ ends of the deletion-spanning contig extended into regions unique to the TBP and CBP, respectively, confirming that, despite the homology surrounding both breakpoints, the deletion-spanning contig indeed spans the GSHPMD deletion (Figure 3C). Taken together, these findings demonstrate that the entirety of the mapped 5.6-Mb region, encompassing the *DMD* and *TMEM47* genes, is deleted from the X chromosome in the GSHPMD model and that the deletion breakpoints are highly homologous to each other.

Identification of Homologous FTHL Loci Present on the Dog X Chromosome

To further investigate the identified homology between the deletion breakpoints, we generated a Pustell matrix comparing the DNA sequence of the deletion-spanning contig and 26- to 33-Mb base pairs of the dog X chromosome, allowing for comparison of DNA in both orientations. This revealed four additional locations of the X chromosome with >90% homology to the deletion-spanning contig, although in opposite orientation to the previously identified deletion breakpoints (Figure 4A).

To determine the identity of the six homologous regions, we approximated their locations on the X chromosome and searched for available annotations within the dog genome assembly via NCBI. Five of the six homologous regions were annotated in the dog genome, and, surprisingly, all five regions were designated as ferritin heavy chain-like (FTHL), two being protein-coding genes and three being pseudogenes (Figure 4B). Importantly, the unidentified, sixth region overlapped with the TBP and extended into the aforementioned assembly gap, likely accounting for its lack of annotation in the dog genome assembly. We found that, while the shared homology spanned nearly the entire length of the pseudogenes, it was unique to only a portion of the protein-coding genes, specifically exon 2 (Figure 4C). Using the encoded peptide sequence of exon 2 of these genes, we queried reference proteomes using HMMER,[19] which uses hidden Markov modeling to search for homologous proteins, and we found that exon 2 is similar to the ferritin-like domain (Data S1). Intriguingly, a DNA sequence alignment of the six homologous regions identified in our Pustell matrix revealed that >96% identities are conserved between the pseudogenes and exon 2 of the protein-coding genes (Figure S7), raising questions as to the mechanism responsible for such high sequence conservation among FTHL genes and pseudogenes alike.

Of the identified ferritin loci, two have NCBI descriptions that link them to the protein-coding, ferritin heavy polypeptide 1 gene (*FTH1*) (Figure 4B). Of note, we identified two copies of *FTH1* in the dog, one with introns located on chromosome 18 and another lacking introns located on chromosome 11 (Figure S3A), and we demonstrated that the cDNA sequences of the dog *FTH1* genes are 100% identical. Further examination revealed a polyadenylation (polyA) signal, AATAAA, followed closely by a string of forty adenine residues, both well-characterized hallmarks of processed mRNAs, uniquely in the intron-lacking copy of the dog *FTH1* gene on chromosome 11 (Figure S3).[20–22] A brief search of NCBI suggested that the additional, intron-lacking copy of *FTH1* is unique to the dog, despite the fact that the intron-containing gene is evolutionarily ancient, as marked by the presence of an ortholog in a wide range of species, including the elephant shark, chicken, cat, and human (data not shown). These findings provide strong evidence to suggest that the intron-lacking copy of *FTH1* on dog chromosome 11 is a processed pseudogene, known to arise through reverse transcription and integration of an mRNA, and, further, that it is of recent origin.

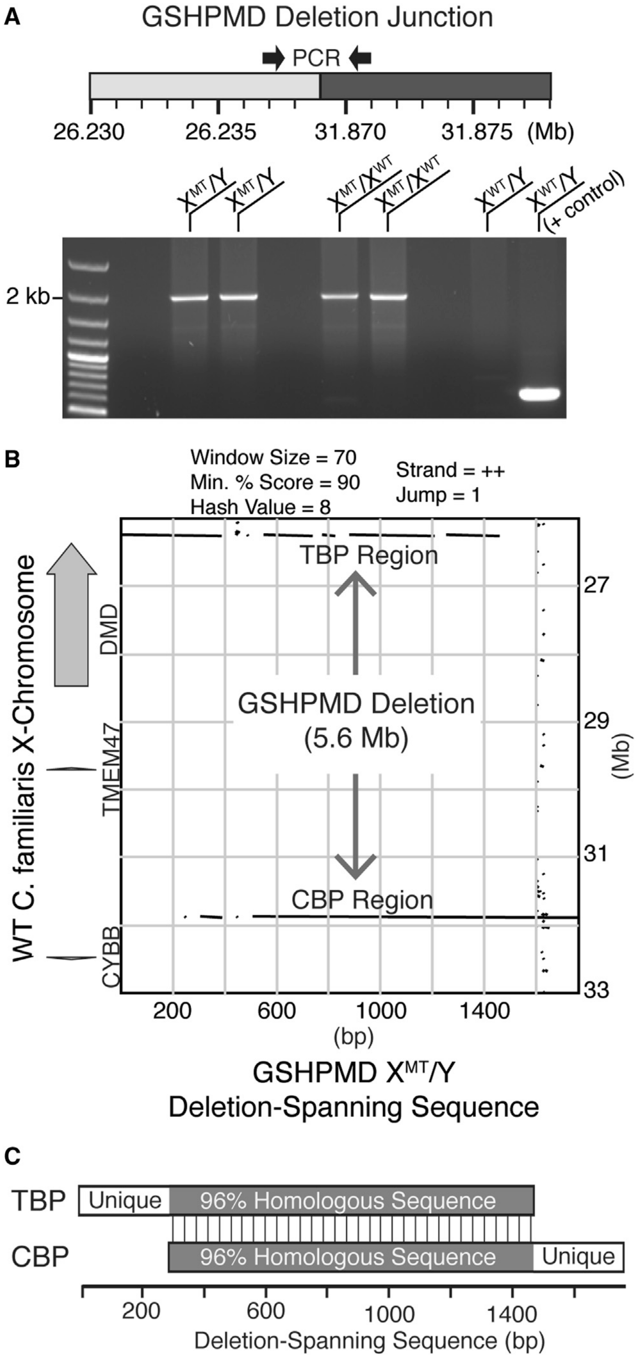


Figure 3. PCR Amplification across the 5.6-Mb Deletion in the GSHPMD Model

(A) Schematic showing primers that are spaced over 5.6 Mb apart in wild-type dog. PCR products generated from this primer pair using DNA from two affected males, two carrier females, and one wild-type male are displayed on an agarose gel following electrophoresis. A positive control using an unrelated primer pair is provided in the rightmost lane for the wild-type dog. (B) Pustell DNA matrix comparing the sequenced deletion-spanning amplicon from a GSHPMD male to the indicated region of wild-type dog X chromosome. Black lines indicate homology between the compared sequences. (C) Macroscale DNA sequence comparison of the telomeric and centromeric deletion breakpoints. Shaded region indicates >96% homology between the deletion breakpoints. X^{MT}/Y, affected male; X^{MT}/X^{WT}, carrier female; X^{WT}/Y, wild-type male; TBP, telomeric breakpoint; CBP, centromeric breakpoint; WT, wild-type.

Next, we questioned whether the identified FTHL pseudogenes (Figure 4B) might actually be processed pseudogenes. We generated a ClustalW sequence alignment of the six homologous regions, but this time we extended the lengths of the individual sequences in the 3′ direction. To our surprise, we identified a polyA signal followed by an A-rich stretch of DNA that varied in length in all six homologous regions, perhaps representing the remnants of what was once a polyA tail but that also could be an unrelated tandem repeat (Figure S2B). We included the second exons and a portion of their 3′ introns from the two protein-coding genes, LOC612257 and FTH1P18, in our alignment, and we found that the polyA signal and A-rich stretch are located just inside the 3′ intron (Figures S3D and S8). While the dog FTH1 processed pseudogene contains a 100% intact polyA signal and tail suggesting it is of recent origin, the identified FTHL genes and pseudogenes contain a less indicative polyA tail, more in line with an earlier origin in evolutionary history, and, therefore, also might be present in other species.

The Identified FTHL Loci Pseudogenes Are Syntenically Conserved among Mammals

The ferritins are evolutionarily ancient iron-binding proteins that are part of the large ferritin-like superfamily. Many copies of the ferritin-H subunit are known to exist as processed pseudogenes in the human.[23–25] Therefore, we asked whether the identified dog FTHL loci (Figure 4B) also are present in other mammals.

To probe this possibility, we used a phylogenetic approach to determine when in evolutionary history the FTHL loci originated. Using the dog FTH1 amino acid sequence as query, we performed a tBLASTn search[26] of mammalian (human, chimpanzee, mouse, dog, cat, and pig), marsupial (opossum), and aves (chicken) genomes. Large numbers of hits were returned for all species, with the exception of the chicken (Figure 5A). To select for true FTHL-processed pseudogenes and intron-lacking genes, we established criteria by which to filter the tBLASTn results (Figure S4). In brief, filtering tBLASTn hits by length and identity criteria substantially reduced the number of hits returned by our search. Interestingly, a comparison of the filtered tBLASTn hits located on the *DMD*-containing chromosome of the queried species revealed that mammals contained a relatively similar numbers of hits, whereas marsupials contained fewer, and aves contained none (Figure 5B).

In addition, a syntenic, multi-species comparison of *DMD* and its surrounding chromosomal region revealed that the locations of the earlier identified dog FTHL loci (Figure 4B) are conserved in mammals, and they can be grouped into four chromosomal regions (Figures 6A and S5A). Although the number of FTHL loci within each of the syntenic chromosomal regions varied by species, our tBLASTn search revealed at least one copy per region in the queried mammals,

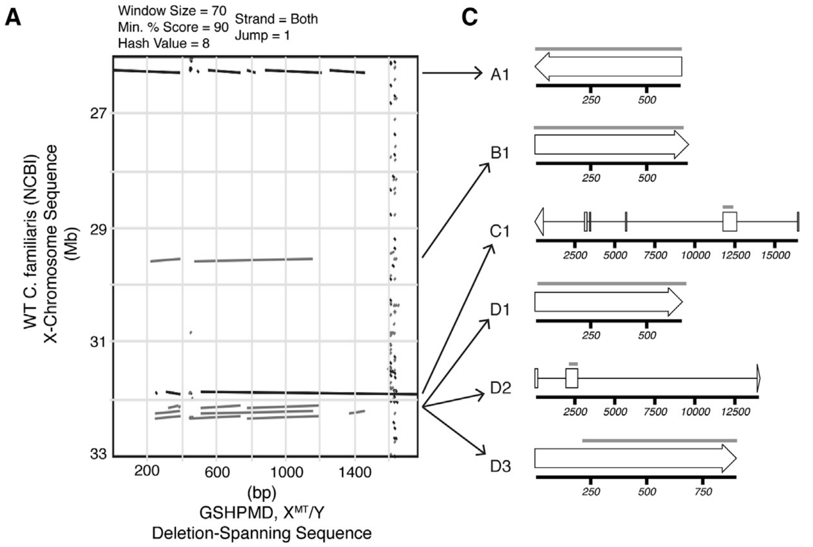


Figure 4. Identification of Homologous DNA Segments on Dog X Chromosome as Members of the Ferritin-like Superfamily

(A) Pustell DNA matrix comparing both orientations of the sequenced deletion-spanning amplicon from a GSHPMD male to the indicated region of wild-type dog X chromosome from the dog reference genome. Strand homology is provided in the subsequent table. (B) Regions of homology identified in Pustell matrix and corresponding gene annotations for these regions of the dog genome from NCBI. Loci are grouped and labeled based on chromosomal location. (C) Schematic of identified ferritin-like genes and pseudogenes. Arrows indicate the location of the genes and pseudogenes in the Pustell matrix. Area of shared homology is shown with a gray bar above each respective locus. *Gene inferred from mammalian FTHL17 ortholog.

B

Pustell Matrix Data		NCBI Gene Information [*Canis lupus familiaris* (dog)]					
Label	Shared Homology	ID	Orientation	Location on X-CHR	Size (bp)	Type	Description
A1	26238481:26239032	FTHL17*	–	N.A.; region is within a genome assembly gap			
B1	29541841:29542399	LOC102153989	+	29541835:29542521	687	pseudogene	ferritin heavy chain pseudogene
C1	31869060:31869618	LOC612257	–	31857202:31873700	16499	protein-coding	ferritin heavy chain-like
D1	32108753:32109313	LOC100687930	+	32108735:32109393	659	pseudogene	ferritin heavy chain pseudogene
D2	32205689:32206259	FTH1P18	+	32203684:32217758	14075	protein-coding	ferritin heavy polypeptide 1, pseudogene 18
D3	32293541:32294101	LOC612281	+	32293314:32294213	900	pseudogene	ferritin heavy polypeptide 1, pseudogene

with the exception of the mouse and the four corresponding syntenic regions of the opossum and chicken (Table S1). In agreement with our tBLASTn search, NCBI designated nearly all of the mammalian hits as FTHL genes or pseudogenes. While not yet formally annotated in the dog genome, our phylogenetic analysis shows that the unidentified dog locus, located in the TBP of the GSHPMD deletion, is sytenically conserved among several mammalian species, suggesting that it is the dog ortholog of the human ferritin heavy polypeptide-like 17 (FTHL17) gene. Furthermore, a multi-species sequence alignment and phylogeny of the FTHL genes and pseudogenes revealed that, of the species examined, the extreme homology among the identified X chromosome FTHL loci is unique to the dog. Interestingly, a subset of the FTHL loci also was highly homologous in the mouse, though this subset was closely linked, spanning less than 85 kb, as compared to the homology shared among the entire set of the identified FTHL loci that span more than 6 Mb in the dog (Figures 6B and S5).

In summary, these findings suggest that the identified FTHL genes and pseudogenes are evolutionarily ancient, arising prior to mammalian radiation. Yet, despite millions of years of evolution, the DNA sequences of these FTHL genes and pseudogenes in the dog, and to a lesser degree the mouse, have remained highly homologous. However, pseudogenes, if free from selective pressure, would be expected to diverge in DNA sequence over time, as is seen in the human and chimpanzee (Figure 6). This intriguing finding suggests a possible function, perhaps at the DNA level, for the identified canine FTHL pseudogenes.

DISCUSSION

In this study, we show that the deletion in the GSHPMD model spans 5.6 Mb of the canine X chromosome and encompasses all known exons of the contiguous *DMD* and *TMEM47* genes. Importantly, the GSHPMD model is expected to be devoid of immunological tolerance and sensitization to dystrophin, due to the complete deletion of its respective gene, and, therefore, it should provide an instrumental pre-clinical model for the prediction of immune responses to gene therapies for DMD. Sequence confirmation of the deletion breakpoint insures that no coding exons of *DMD* remain, precluding expression of any dystrophin-derived peptide sequences through exon skipping or alternative transcription from any known internal promoters. The same considerations apply to the contiguous *TMEM47* gene, which encodes a highly conserved 19.9-kDa protein without any identified Mendelian disease association. TMEM47, also known as brain cell membrane protein I, is abundantly transcribed in dog brain.[27] To the best of our knowledge, TMEM47 protein expression has not been characterized in the dog; we were unable to detect TMEM47 protein in brain tissue of wild-type dog by immunohistochemistry or western blot, using a commercially available polyclonal antibody raised against a shared human epitope, and there are no canine-reactive antibodies commercially available (data not shown). TMEM47's candidacy for a role in human X-linked mental retardation has been explored and dismissed.[28] Moreover, the phenotypic difference between dystrophic and normal littermates in the GSHPMD colony appears to be strictly related to muscle impairment, as is the case in the GRMD colony.

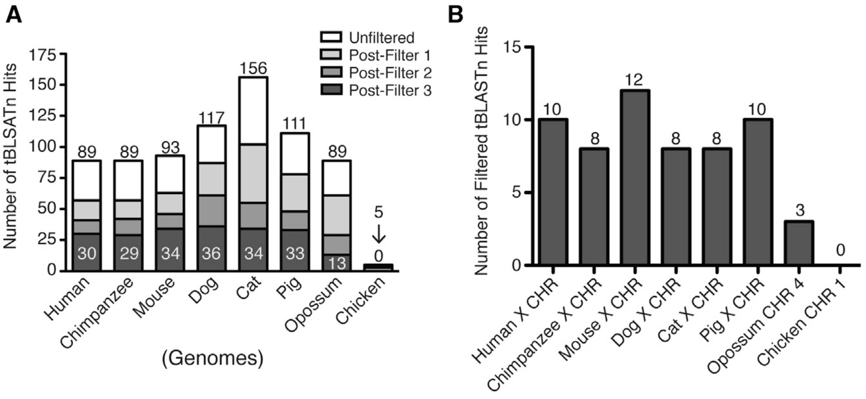


Figure 5. Quantification of Results from tBLASTn Search of Dog FTH1 Peptide in Multiple Species

(A) Bar graph quantifying tBLASTn search results of dog FTH1 peptide sequence against the genomes of the indicated species. Number at the top of each bar signifies hits returned prior to applying filtering criteria. Number internal to bottom section of each bar represents hits remaining after applying all filtering criteria. (B) Bar graph quantifies the number of filtered tBLASTn hits present on the chromosome that harbors the *DMD* gene for the indicated species. CHR, chromosome.

The deletion breakpoints were within highly homologous segments of DNA that NCBI identifies as members of the ferritin family. We uncovered additional copies of the homologous ferritin DNA loci on the dog X chromosome, designated as a mix of protein-coding genes and pseudogenes. These ferritin-like loci are evolutionarily ancient, arising prior to mammalian radiation; but, interestingly, the high homology is unique to the ferritin loci in the dog and, to a lesser degree, the mouse. The observed sequence homology among the identified ferritin genes and pseudogenes is maintained across large genetic distances (>6 Mb) uniquely in the dog, a phenomenon likely arising through gene conversion (reviewed in Chen et al.[29]). These findings suggest that the GSHPMD deletion occurred through a homologous recombination event, an outcome that was likely enabled by gene conversion of these ferritin loci in the canine lineage.

To map the deletion in the GSHPMD model, we used a PCR strategy and found no amplification from primers internal to the deletion. Furthermore, amplification across the 5.6-Mb deletion generated an amplicon uniquely in the GSHPMD model (Figures 2 and 3A), the length and sequence of which demonstrate a clean breakpoint without insertion or rearrangement. This, in combination with data showing that skeletal muscle of GSHPMD dogs lacks dystrophin,[11] provides clear evidence supporting the deletion, rather than translocation, of the 5.6-Mb region of X chromosome that encompasses the *DMD* and *TMEM47* genes. The deletion breakpoints are within highly homologous segments of DNA, which span over 1 kb and share 96% identity (Figure 3C), suggesting that the deletion occurred through a homologous recombination event. Supporting this claim, the region encompassing the TBP of the GSHPMD deletion is a recombination hotspot in the dog genome.[30] Consistent with recombination hotspots in the dog,[30] the GSHPMD deletion breakpoints have >67% GC content over a length of about 550 bp (Figure S7). Furthermore, the *FTHL17* gene, which overlaps with the TBP of the GSHPMD model, has been implicated in the X chromosome translocation breakpoint of an infertile human male.[31] Together, this indirect evidence strongly supports the notion that the GSHPMD deletion likely occurred through homologous recombination.

Next, we discovered that the deletion breakpoints were not only homologous to each other but also homologous to four additional regions of the dog X chromosome. These six regions share >96% identity over a length of about 550 bp, are distributed across 6 Mb of the X chromosome, and are members of the ferritin family (Figures 4 and S7). Interestingly, NCBI designates these ferritin loci as a mixture of protein-coding genes and pseudogenes. However, the high level of homology between the FTHL loci in the dog and a subset in the mouse suggests selective pressure on these loci, perhaps at the DNA level (Figures 6 and S5). Of note, a comparison of the FTHL hits returned by tBLASTn in the human and chimpanzee, which shared a common ancestor less than 10 million years ago,[17] shows that interspecies orthologs are more closely related than intraspecies paralogs (Figure 6B), the expected outcome of divergent evolution. In contrast, the identified FTHL loci in the dog have not diverged and appear to be evolving in concert.

Concerted evolution describes a phenomenon in which individual genetic copies of a multigene family, generally present in a tandem array, evolve in concert; that is, a point mutation in one copy becomes propagated throughout the array of genes (reviewed in Nei and Rooney[32]). This phenomenon was first described during a study of the rRNA genes in the African toad[33] and is thought to occur by gene conversion,[29] a mechanism that utilizes DNA repair machinery to transfer base-pair mismatches between highly homologous yet distinct DNA strands following a double-stranded break. However, in contrast to the ∼450 repeat copies in rRNA genes of the African toad, we found only six highly homologous copies of the FTHL loci that span over 6 Mb on the canine X chromosome. Unequal crossover of the *FTHL17* gene with any of the other identified FTHL loci would result in a large *DMD*-encompassing deletion, such as that seen in the GSHPMD model. However, given that the DMD phenotype is not commonly reported in dogs, the more recently proposed synthesis-dependent strand annealing (SDSA) model,[34] which results in gene conversion yet non-crossover products, is a more plausible mechanism to account for the concerted evolution of the identified FTHL loci. Gene conversion is known to occur across long genetic distances in humans, as with von Willebrand disease, where mutation in the VWD gene can result from interchromosomal

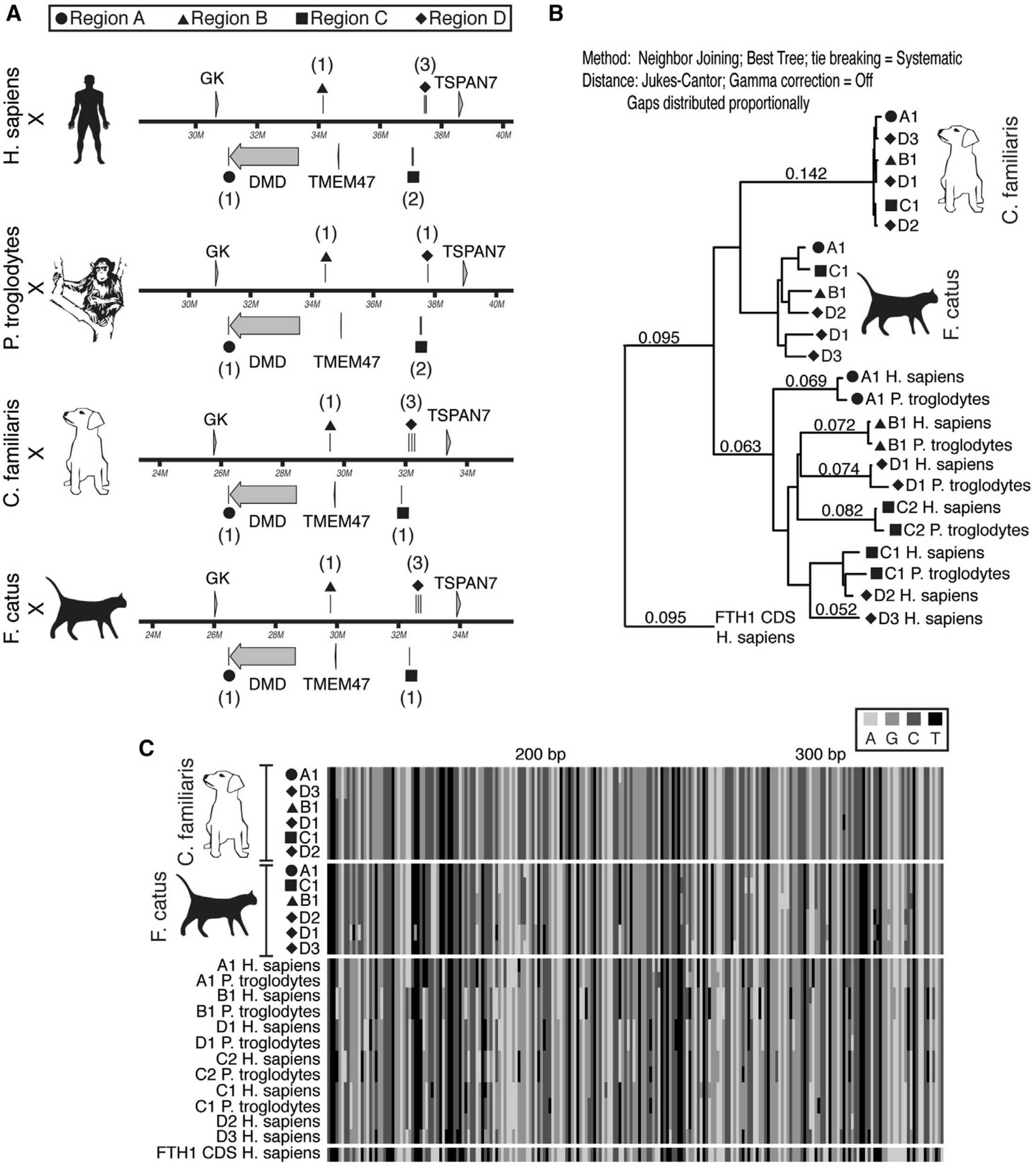


Figure 6. Phylogenetic Analysis of Identified FTHL Loci from tBLASTn Search in Human, Chimpanzee, Dog, and Cat

(A) Comparison of syntenic portion of X chromosome from each species showing the grouping of FTHL loci to four regions, labeled A, B, C, and D. Number of pseudogenes present in reach region is indicated in parentheses. (B) Phylogeny of identified FTHL loci. The human FTH1 CDS is used as an outgroup. Branches with values less than 0.05 are not displayed. Tree-building parameters are provided as text in figure. CDS, coding DNA sequence. (C) ClustalW multiple DNA sequence alignment of identified FTHL loci. Note extreme sequence homology among dog FTHL loci.

gene conversion between the true VWD gene and a highly homologous pseudogene.[35]

While arguing against the intrinsic definition of a pseudogene, we thought the identified FTHL pseudogenes might perform a function in the dog and would, therefore, be conserved due to selective pressure. Indeed, while most pseudogenes lose their transcription potential, those that are transcribed are capable of acting as short interfering RNAs and microRNAs.[36] But, hindering our analysis, the *FTHL17* gene has testis-specific activity in both the human and mouse,[37,38] although the human FTHL17 protein is unstable and has no ferroxidase activity.[39] Perhaps unsurprisingly, a BLASTn[40] search of the dog transcriptome[41] using a FTHL sequence as query resulted in greater than 1,000 hits in testis and less than 10 hits in all other tissue types that were available (data not shown). However, due to the short, 100-bp sequence reads of the dog transcriptome and the extreme homology between the FTHL loci, we were not able to discern whether the observed hits were transcripts resulting from the FTHL genes or the FTHL pseudogenes. Dedicated study of canine FTHL pseudogenes may reveal a function as an RNA species.

Here we provide a detailed characterization of the *DMD* deletion in the GSHPMD model. Our data strongly suggest the deletion arose through homologous recombination of FTHL loci that appear to be evolving in concert in the dog. Due to the complete deletion of the *DMD* gene, this model is expected to be devoid of central immunological tolerance to any portion of the 427-kDa dystrophin protein. Furthermore, this deletion precludes any sensitization to truncated dystrophin peptides that might, with other mutations, arise somatically within revertant fibers. These features of the GSHPMD model favor its use in the rigorous study of cytotoxic immune responses to recombinant dystrophin expressed following regional or systemic gene therapy. Interestingly, intra-breed allelic diversity, while limited, has been identified within DLA-88, a dog leukocyte antigen (DLA) gene that displays the highest polymorphism of the MHC class I loci.[42–44] Therefore, immunological studies within a single dog breed, such as the GSHPMD model, may result in varied responses based on DLA class I haplotype, a result that can be expected in clinical trials due to the highly polymorphic HLAs encoding MHC I.[45] In conclusion, we envision the GSHPMD model will be instrumental in the evaluation of potential immune responses to gene therapies for DMD, which will aid in the demonstration of safety, a matter of paramount importance in the FDA's review of investigational new drug applications for clinical trials.

MATERIALS AND METHODS

Animals

The GSHPMD colony was initially housed at the University of North Carolina at Chapel Hill before being moved to Texas A&M University. Dogs were cared for and assessed according to principles outlined in the National Research Council Guide for the Care and Use of Laboratory Animals and covered by the UNC-CH Institutional Animal Care and Use Committee (IACUC) through protocols, Natural History and Immunological Parameters in the German Shorthaired Pointer Muscular Dystrophy (GSHPMD) Dog (UNC 09-011) and Standard Operating Procedures—Canine X-Linked Muscular Dystrophy (UNC 09-351 and TAMU IACUC 2015-0110).

Primer Design

Detailed primer information is available in Figure S9. Primers were designed using MacVector v14.5.0 software, and then they were checked for specificity to the region of interest of the dog genome using Primer-Blast.[46] Primers were ordered from Integrated DNA Technologies. Annealing temperatures were calculated as five degrees less than the average melting temperature of the primers in a reaction. Melting temperatures were calculated using OligoAnalyzer v3.1 (IDT). Only primers that yielded the expected band size when amplified from WT dog DNA were used in our analysis.

PCR in Deletion Mapping

Genomic DNA was isolated from dog whole blood using the QIAamp DNA Blood Midi Kit (QIAGEN) following the manufacturer's protocol. PCRs contained 0.5 μM of each primer, 200 ng genomic DNA, and 25 μL GoTaq Green Master Mix (Promega), and they were brought to a final volume of 50 μL. PCRs were performed on a PTC-200 DNA Engine thermocycler (MJ Research). Reactions were initially denatured at 95°C for 2 min and then carried through 35 cycles as follows: 95°C for 45 s, annealing for 45 s, extension at 72°C, followed by a final extension at 72°C for 10 min, and then held indefinitely at 4°C. Extension times were calculated using the processivity of standard Taq polymerase, 1 kb per minute, but reactions were extended for a minimum of 30 s. Completed PCR reactions were run on a 0.5% agarose gel, stained with ethidium bromide (Sigma-Aldrich), and captured on a digital imager (Fotodyne). PCR products from either male GSHPMD DNA or WT male GSHP DNA were compared in this manner for each primer pair.

In the case of amplifying across the deletion, reactions contained 0.5 μM primer 145 and 146, 1.5 M betaine (Sigma-Aldrich), 200 ng genomic DNA, and 25 μL GoTaq Green Master Mix (Promega), and they were brought to a final volume of 50 μL. Reactions were run on a Multigene Gradient thermocycler (Labnet International) under the following conditions: 95°C for 2 min, 35 cycles of 95°C for 30 s, 64.5°C for 30 s, 72°C for 7 min, followed by a final extension at 72°C for 10 min, and then held indefinitely at 4°C. Completed reactions were run on a 0.5% agarose gel by electrophoresis. To avoid UV-induced mutations prior to sequencing, replicate lanes were run, cut away from the gel, and stained with ethidium bromide. The position of the 2-kb band was located on the ethidium bromide-stained lanes, and then it was used to approximate the location of the band of interest on the unstained, unexposed lanes, which was excised and purified using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's protocol. The resulting DNA was used in subsequent sequencing reactions.

Assembly Gap PCR

BAC clone CH82-472H14 was ordered from the BACPAC Resource Center (CHORI). The BAC clone was provided to us in DH10B *E. coli*, which we propagated in LB broth (Corning Life Sciences) with 12.5 μg/mL chloramphenicol in an incubator at 37°C and 220 rpm for 20 hr. BAC DNA was purified from *E. coli* using the Plasmid Mini Kit (QIAGEN) following the manufacturer's protocol.

PCRs contained 0.5 μM primer 147 and 148, 1.5 M betaine (Sigma), 5 pg BAC DNA, and 25 μL GoTaq Green Master Mix (Promega), and they were brought to a final volume of 50 μL. PCRs were performed on a PTC-200 DNA Engine thermocycler (MJ Research). Reactions were initially denatured at 95°C for 2 min and then carried through 35 cycles as follows: 95°C for 30 s, 58°C for 30 s, 72°C for 2 min, followed by a final extension at 72°C for 10 min, and then held indefinitely at 4°C. Completed reactions were run on a 0.8% agarose gel by electrophoresis, and the expected 1-kb band was excised without

UV exposure or ethidium bromide staining (described in previous section) and then purified using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's protocol. The resulting DNA was used in subsequent sequencing reactions.

DNA Sequencing

The assembly gap and deletion-spanning amplicons were sequenced by the DNA Sequencing Facility at The University of Pennsylvania on a 3730xl DNA Analyzer using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) according to the kit instructions. Sequencing reactions contained 8 μL Terminator Mix, 0.2 μM primer, and 100 ng gel-purified DNA in a final volume of 20 μL. Individual Sanger reads were manually trimmed, as we found software-directed trimming to be too lenient in eliminating poor-quality base calls. Trimmed reads were then assembled into a single contig using SeqMan Pro v12.0.0 (DNASTAR) under default settings. Individual Sanger reads used in our contig assembles are available in Trace Archive under the following sequential accession numbers: GSHPMD deletion reads 2,342,817,208–2,342,817,224; assembly gap reads 2,342,817,194–2,342,817,207.

Pustell Matrices, ClustalW Sequence Alignments, and Phylogenies

Pustell DNA matrices, ClustalW[47] alignments, and phylogenies were created in MacVector. Pustell matrices were scored with the default DNA database matrix provided by MacVector. Specific parameters for matrices and phylogenies are provided in their respective figures. ClustalW multiple sequence alignments were performed under MacVector default conditions, which include an open gap penalty of 15.0, extend gap penalty of 6.7, and delay divergence of 30%. DNA sequences used in the alignments can be retrieved from NCBI using the coordinates provided in Figure 4B and Table S1.

tBLASTn Search

A tBLASTn search[26] was performed under default conditions using the dog FTH1 peptide sequence as query against many species' genome assemblies. Search results were exported as.csv files, which were then filtered by length and quality parameters (Figure S4). Individual hit information for each species is provided in Data S2.

Genome Assembly Builds

The following genome assembly builds were accessed via NCBI and used in our analysis (the described species are organized with common name, *scientific name*, genome assembly build): dog, *Canis lupus familiaris*, CanFam3.1; human, *Homo sapiens*, GRCh38.p3; chimpanzee, *Pan troglodytes*, Pan_troglodytes-2.1.4; mouse, *Mus musculus*, GRCm38.p4; cat, *Felis catus*, Felis_catus_8.0; pig, *Sus scrofa*, Sscrofa10.2; opossum, *Monodelphis domestica*, MonDom5; and chicken, *Gallus gallus*, Gallus_gallus-4.0. It should be noted that much of the gene and locus information provided herein is predicted model sequences produced by NCBI's eukaryotic genome annotation pipeline and, therefore, subject to change due to the dynamic nature of genome assemblies and annotation software.[48]

SUPPLEMENTAL INFORMATION

Supplemental Information includes nine figures, one table, and two data files and can be found with this article online at http://dx.doi.org/10.1016/j.omtm.2016.12.001.

AUTHOR CONTRIBUTIONS

Conceptualization, D.J.V., P.S.H., J.N.K., H.H.S.; Methodology, D.J.V., J.N.K., H.H.S.; Investigation, D.J.V., A.S.M.; Resources, J.N.K.; Writing – Original Draft, D.J.V.; Writing – Review and Editing, A.S.M., P.S.H., J.N.K., H.H.S.; Visualization, D.J.V., A.S.M., J.N.K., H.H.S.; Supervision, P.S.H., H.H.S.; Funding Acquisition, J.N.K., H.H.S.

ACKNOWLEDGMENTS

This work was supported by The National Institute of Neurological Disorders and Stroke (R01NS094705 and R01NS042874 to H.H.S.), The National Center for Research Resources (S10RR028027 to H.H.S.), The National Institute of Arthritis and Musculoskeletal and Skin Diseases (P30AR050950 and T32AR053461), and The Muscular Dystrophy Association (276601 to H.H.S. and 86447 to J.N.K.).

REFERENCES

1. Ramos, J., and Chamberlain, J.S. (2015). Gene Therapy for Duchenne muscular dystrophy. Expert Opin. Orphan Drugs *3*, 1255–1266.
2. Potter, M.A., and Chang, P.L. (1999). Review–the use of immunosuppressive agents to prevent neutralizing antibodies against a transgene product. Ann. N Y Acad. Sci. *875*, 159–174.
3. Yang, Y., Jooss, K.U., Su, Q., Ertl, H.C., and Wilson, J.M. (1996). Immune responses to viral antigens versus transgene product in the elimination of recombinant adenovirus-infected hepatocytes in vivo. Gene Ther. *3*, 137–144.
4. Sicinski, P., Geng, Y., Ryder-Cook, A.S., Barnard, E.A., Darlison, M.G., and Barnard, P.J. (1989). The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science *244*, 1578–1580.
5. Sharp, N.J., Kornegay, J.N., Van Camp, S.D., Herbstreith, M.H., Secore, S.L., Kettle, S., Hung, W.Y., Constantinou, C.D., Dykstra, M.J., Roses, A.D., et al. (1992). An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics *13*, 115–121.
6. Hoffman, E.P., Morgan, J.E., Watkins, S.C., and Partridge, T.A. (1990). Somatic reversion/suppression of the mouse mdx phenotype in vivo. J. Neurol. Sci. *99*, 9–25.
7. Pigozzo, S.R., Da Re, L., Romualdi, C., Mazzara, P.G., Galletta, E., Fletcher, S., Wilton, S.D., and Vitiello, L. (2013). Revertant fibers in the mdx murine model of Duchenne muscular dystrophy: an age- and muscle-related reappraisal. PLoS ONE *8*, e72147.
8. Schatzberg, S.J., Anderson, L.V., Wilton, S.D., Kornegay, J.N., Mann, C.J., Solomon, G.G., and Sharp, N.J. (1998). Alternative dystrophin gene transcripts in golden retriever muscular dystrophy. Muscle Nerve *21*, 991–998.
9. Flanigan, K.M., Campbell, K., Viollet, L., Wang, W., Gomez, A.M., Walker, C.M., and Mendell, J.R. (2013). Anti-dystrophin T cell responses in Duchenne muscular dystrophy: prevalence and a glucocorticoid treatment effect. Hum. Gene Ther. *24*, 797–806.
10. Olby, N.J., Sharp, N.J., Nghiem, P.E., Keene, B.W., DeFrancesco, T.C., Sidley, J.A., Kornegay, J.N., and Schatzberg, S.J. (2011). Clinical progression of X-linked muscular dystrophy in two German Shorthaired Pointers. J. Am. Vet. Med. Assoc. *238*, 207–212.
11. Schatzberg, S.J., Olby, N.J., Breen, M., Anderson, L.V., Langford, C.F., Dickens, H.F., Wilton, S.D., Zeiss, C.J., Binns, M.M., Kornegay, J.N., et al. (1999). Molecular analysis of a spontaneous dystrophin 'knockout' dog. Neuromuscul. Disord. *9*, 289–295.

12. Brown, J., Dry, K.L., Edgar, A.J., Pryde, F.E., Hardwick, L.J., Aldred, M.A., Lester, D.H., Boyle, S., Kaplan, J., Dufier, J.L., et al. (1996). Analysis of three deletion breakpoints in Xp21.1 and the further localization of RP3. Genomics *37*, 200–210.

13. Francke, U., Ochs, H.D., de Martinville, B., Giacalone, J., Lindgren, V., Distèche, C., Pagon, R.A., Hofker, M.H., van Ommen, G.J., Pearson, P.L., et al. (1985). Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome. Am. J. Hum. Genet. *37*, 250–267.

14. Kunkel, L.M., Monaco, A.P., Middlesworth, W., Ochs, H.D., and Latt, S.A. (1985). Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. Proc. Natl. Acad. Sci. USA *82*, 4778–4782.

15. Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertelson, C.J., Kurnit, D.M., and Kunkel, L.M. (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature *323*, 646–650.

16. Smith, T.J., Wilson, L., Kenwrick, S.J., Forrest, S.M., Speer, A., Coutelle, C., and Davies, K.E. (1987). Isolation of a conserved sequence deleted in Duchenne muscular dystrophy patients. Nucleic Acids Res. *15*, 2167–2174.

17. Benton, M.J., and Donoghue, P.C. (2007). Paleontological evidence to date the tree of life. Mol. Biol. Evol. *24*, 26–53.

18. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803–819.

19. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.

20. Fitzgerald, M., and Shenk, T. (1981). The sequence 5'-AAUAAA-3'forms parts of the recognition site for polyadenylation of late SV40 mRNAs. Cell *24*, 251–260.

21. Proudfoot, N.J., and Brownlee, G.G. (1976). 3′ non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211–214.

22. Tabaska, J.E., and Zhang, M.Q. (1999). Detection of polyadenylation signals in human DNA sequences. Gene *231*, 77–86.

23. Andrews, S.C. (2010). The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. Biochim. Biophys. Acta *1800*, 691–705.

24. Costanzo, F., Colombo, M., Staempfli, S., Santoro, C., Marone, M., Frank, R., Delius, H., and Cortese, R. (1986). Structure of gene and pseudogenes of human apoferritin H. Nucleic Acids Res. *14*, 721–736.

25. Zhang, Z., Harrison, P.M., Liu, Y., and Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. *13*, 2541–2558.

26. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

27. Christophe-Hobertus, C., Szpirer, C., Guyon, R., and Christophe, D. (2001). Identification of the gene encoding Brain Cell Membrane Protein 1 (BCMP1), a putative four-transmembrane protein distantly related to the Peripheral Myelin Protein 22 / Epithelial Membrane Proteins and the Claudins. BMC Genomics *2*, 3.

28. Christophe-Hobertus, C., Kooy, F., Gecz, J., Abramowicz, M.J., Holinski-Feder, E., Schwartz, C., and Christophe, D. (2004). TM4SF10 gene sequencing in XLMR patients identifies common polymorphisms but no disease-associated mutation. BMC Med. Genet. *5*, 22.

29. Chen, J.M., Cooper, D.N., Chuzhanova, N., Férec, C., and Patrinos, G.P. (2007). Gene conversion: mechanisms, evolution and human disease. Nat. Rev. Genet. *8*, 762–775.

30. Axelsson, E., Webster, M.T., Ratnakumar, A., Ponting, C.P., and Lindblad-Toh, K.; LUPA Consortium (2012). Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res. *22*, 51–63.

31. Lee, S., Lee, S.H., Chung, T.G., Kim, H.J., Yoon, T.K., Kwak, I.P., Park, S.H., Cha, W.T., Cho, S.W., and Cha, K.Y. (2003). Molecular and cytogenetic characterization of two azoospermic patients with X-autosome translocation. J. Assist. Reprod. Genet. *20*, 385–389.

32. Nei, M., and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. *39*, 121–152.

33. Brown, D.D., Wensink, P.C., and Jordan, E. (1972). A comparison of the ribosomal DNA's of Xenopus laevis and Xenopus mulleri: the evolution of tandem genes. J. Mol. Biol. *63*, 57–73.

34. Haber, J.E., Ira, G., Malkova, A., and Sugawara, N. (2004). Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. Philos. Trans. R. Soc. Lond. B Biol. Sci. *359*, 79–86.

35. Gupta, P.K., Adamtziki, E., Budde, U., Jaiprakash, M., Kumar, H., Harbeck-Seu, A., Kannan, M., Oyen, F., Obser, T., Wedekind, I., et al. (2005). Gene conversions are a common cause of von Willebrand disease. Br. J. Haematol. *130*, 752–758.

36. Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., and Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA *17*, 792–798.

37. Kobayashi, S., Fujihara, Y., Mise, N., Kaseda, K., Abe, K., Ishino, F., and Okabe, M. (2010). The X-linked imprinted gene family Fthl17 shows predominantly female expression following the two-cell stage in mouse embryos. Nucleic Acids Res. *38*, 3672–3681.

38. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes expressed in spermatogonia. Nat. Genet. *27*, 422–426.

39. Ruzzenenti, P., Asperti, M., Mitola, S., Crescini, E., Maccarinelli, F., Gryzik, M., Regoni, M., Finazzi, D., Arosio, P., and Poli, M. (2015). The Ferritin-Heavy-Polypeptide-Like-17 (FTHL17) gene encodes a ferritin with low stability and no ferroxidase activity and with a partial nuclear localization. Biochim. Biophys. Acta *1850*, 1267–1273.

40. Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. J. Comput. Biol. *7*, 203–214.

41. Hoeppner, M.P., Lundquist, A., Pirun, M., Meadows, J.R., Zamani, N., Johnson, J., Sundström, G., Cook, A., FitzGerald, M.G., Swofford, R., et al. (2014). An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. PLoS ONE *9*, e91172.

42. Graumann, M.B., DeRose, S.A., Ostrander, E.A., and Storb, R. (1998). Polymorphism analysis of four canine MHC class I genes. Tissue Antigens *51*, 374–381.

43. Kennedy, L.J., Barnes, A., Happ, G.M., Quinnell, R.J., Courtenay, O., Carter, S.D., Ollier, W.E., and Thomson, W. (2002). Evidence for extensive DLA polymorphism in different dog populations. Tissue Antigens *60*, 43–52.

44. Ross, P., Buntzman, A.S., Vincent, B.G., Grover, E.N., Gojanovich, G.S., Collins, E.J., Frelinger, J.A., and Hess, P.R. (2012). Allelic diversity at the DLA-88 locus in Golden Retriever and Boxer breeds is limited. Tissue Antigens *80*, 175–183.

45. Manno, C.S., Pierce, G.F., Arruda, V.R., Glader, B., Ragni, M., Rasko, J.J., Ozelo, M.C., Hoots, K., Blatt, P., Konkle, B., et al. (2006). Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat. Med. *12*, 342–347.

46. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.

47. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

48. Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. *40*, D130–D135.

OMTM, Volume *4*

Supplemental Information

Mechanism of Deletion Removing All Dystrophin Exons in a Canine Model for DMD Implicates Concerted Evolution of X Chromosome Pseudogenes

D. Jake VanBelzen, Alock S. Malik, Paula S. Henthorn, Joe N. Kornegay, and Hansell H. Stedman

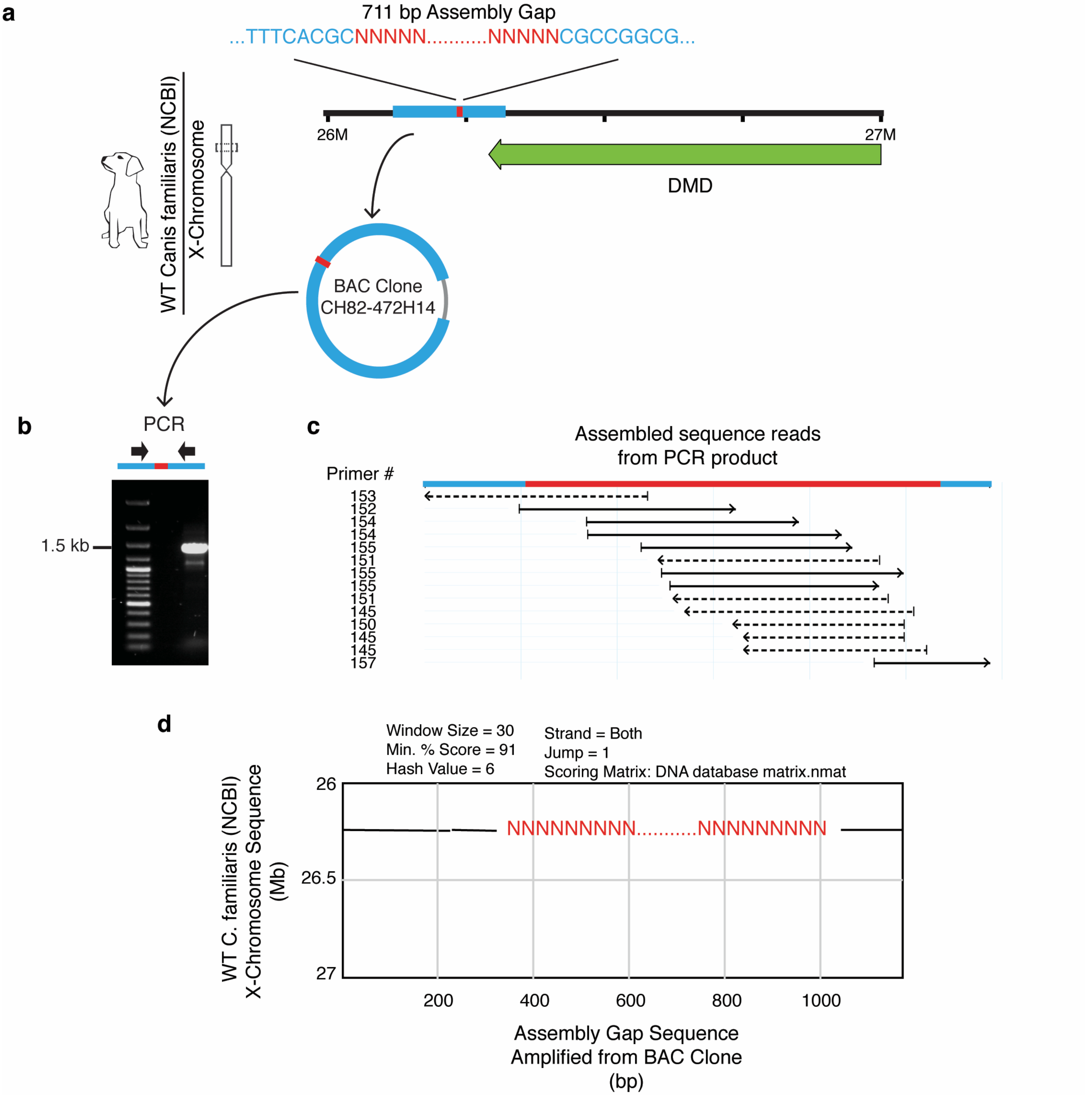
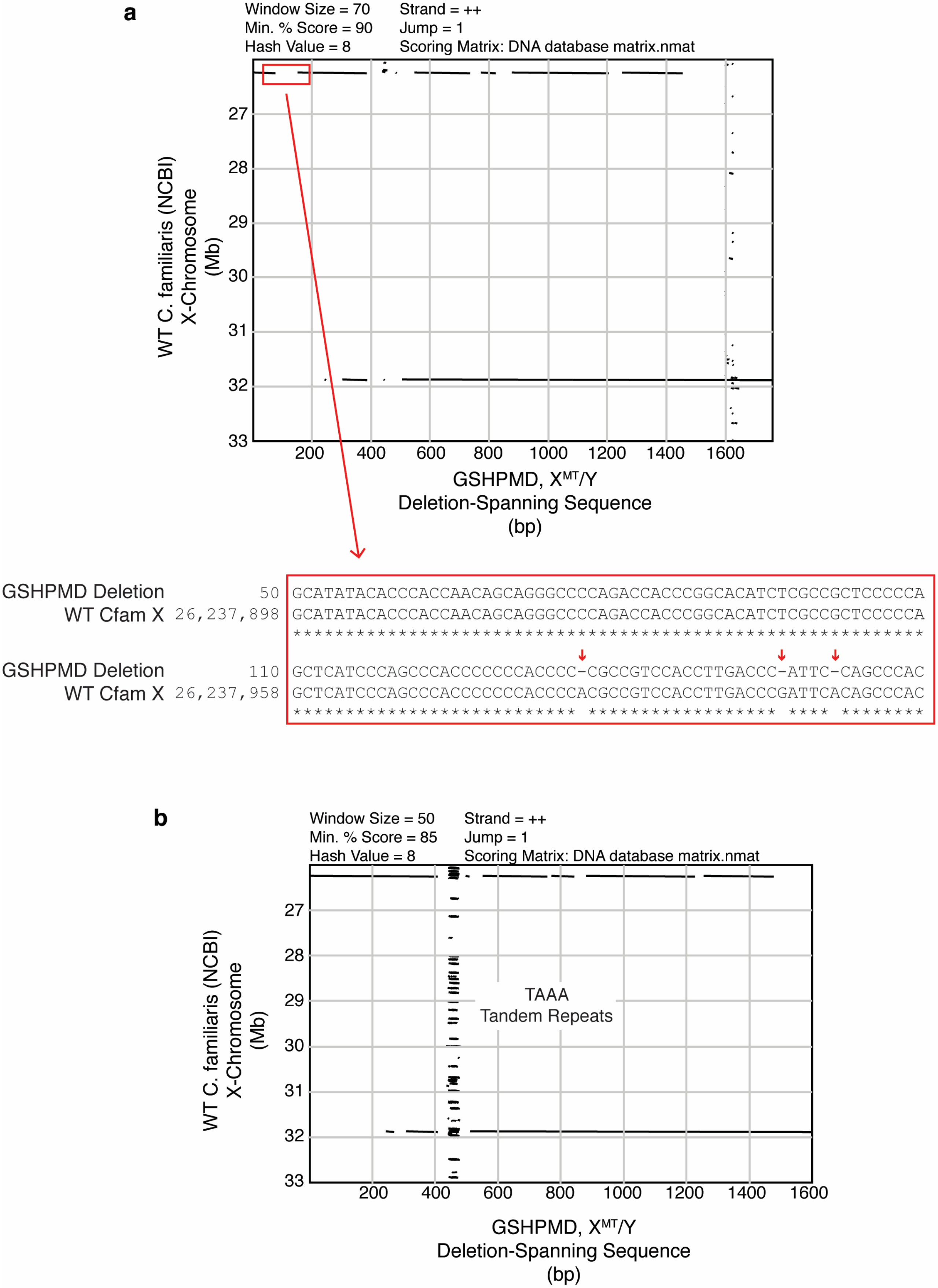


Figure S1. Sequencing of assembly gap in dog reference genome near telomeric breakpoint of GSHPMD deletion. (**a**) A region of the reference dog X-chromosome is shown. The region of the X-chromosome contained within the library BAC clone is highlighted in blue. The assembly gap is shown in red. BAC, bacterial artificial chromosome. N, any nucleotide. (**b**) Primers, depicted by arrows, that anneal outside opposite ends of the assembly gap were used to PCR amplify across the assembly gap using the BAC clone as template. PCR products are displayed on an agarose gel following electrophoresis. (**c**) The gel-purified PCR product was Sanger sequenced using the indicated primers, and overlapping reads were assembled into a single contig. (**d**) Pustell matrix comparing the sequenced PCR product to the indicated region of the dog X-chromosome reference sequence. Internal break in homology is expected and represents the assembly gap in the dog reference sequence, which is depicted by a stretch of red N’s.



```
GSHPMD Deletion         50 GCATATACACCCACCAACAGCAGGGCCCCAGACCACCCGGCACATCTCGCCGCTCCCCCA
     WT Cfam X  26,237,898 GCATATACACCCACCAACAGCAGGGCCCCAGACCACCCGGCACATCTCGCCGCTCCCCCA
                           ************************************************************

GSHPMD Deletion        110 GCTCATCCCAGCCCACCCCCCCACCCC-CGCCGTCCACCTTGACCC-ATTC-CAGCCCAC
     WT Cfam X  26,237,958 GCTCATCCCAGCCCACCCCCCCACCCCACGCCGTCCACCTTGACCCGATTCACAGCCCAC
                           *************************** ****************** **** ********
```

Figure S2. Investigation of breaks in homology between GSHPMD deletion-spanning sequence and dog genome reference sequence. (**a**) Pustell matrix comparing deletion-spanning sequence from GSHPMD to a region of the dog X-chromosome reference sequence. Red box highlights a break in sequence homology near the telomeric breakpoint, and a ClustalW alignment of the corresponding sequences is shown below. Red arrows indicate three indels responsible for the observed lapse in homology. (**b**) Pustell matrix, with reduced specificity parameters, depicting the presence of a TAAA tandem repeat present in this region of the dog X-chromosome.

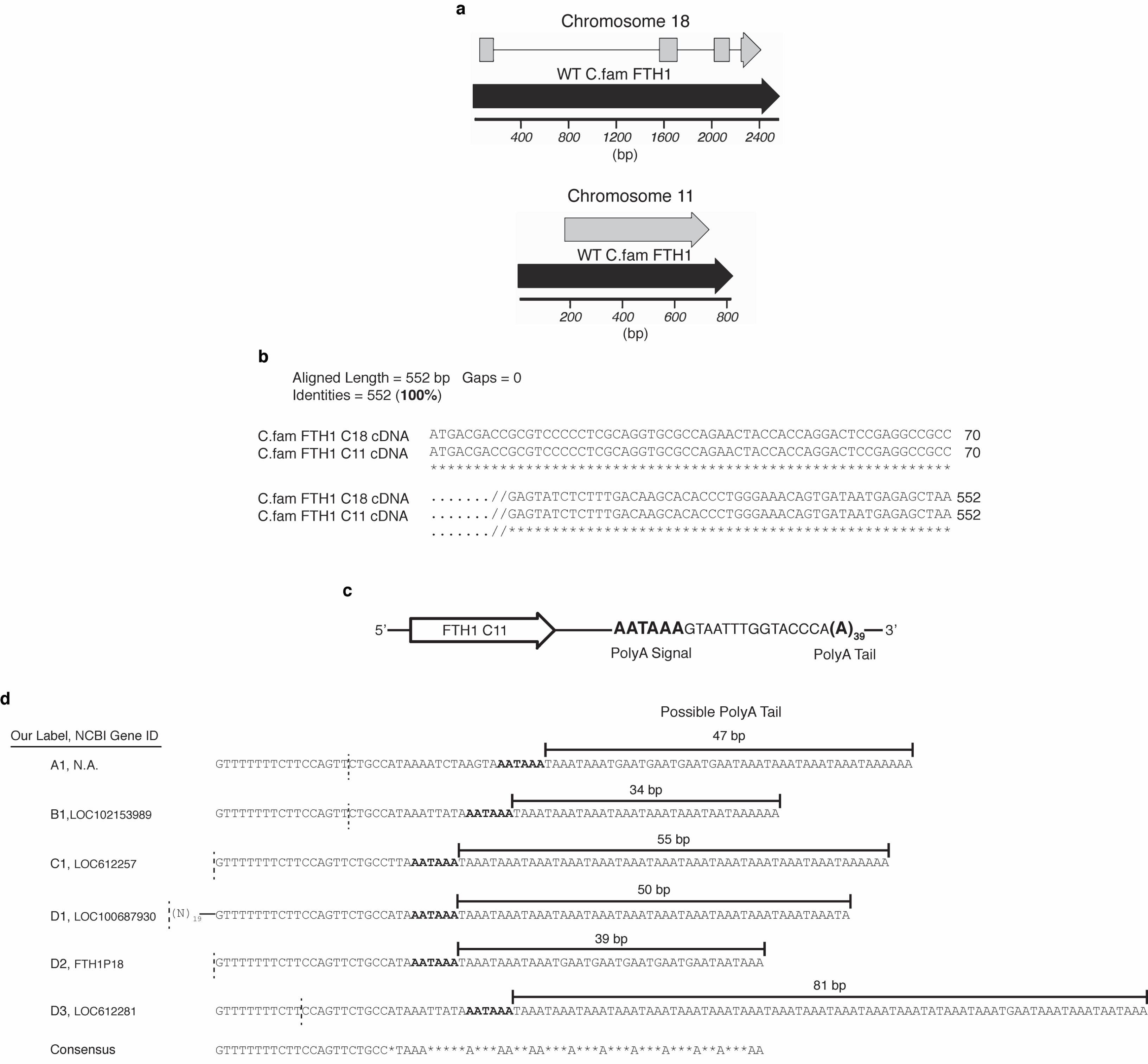
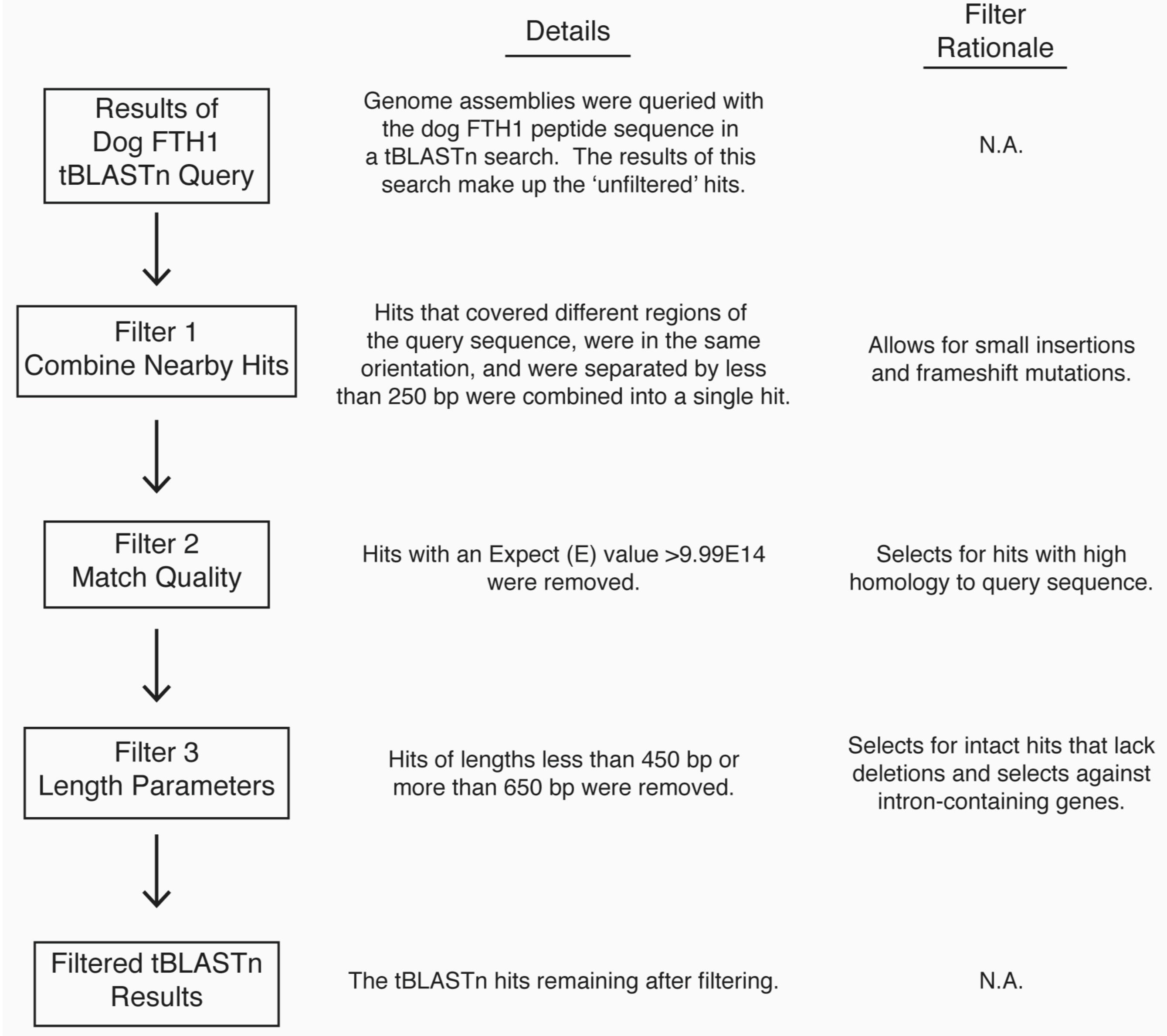

a
Chromosome 18
WT C.fam FTH1
400
800
1200
1600
2000
2400
(bp)
Chromosome 11
WT C.fam FTH1
200
400
600
800
(bp)
b
Aligned Length = 552 bp Gaps = 0
Identities = 552 (**100%**)
C.fam FTH1 C18 cDNA ATGACGACCGCGTCCCCCTCGCAGGTGCGCCAGAACTACCACCAGGACTCCGAGGCCGCC 70
C.fam FTH1 C11 cDNA ATGACGACCGCGTCCCCCTCGCAGGTGCGCCAGAACTACCACCAGGACTCCGAGGCCGCC 70
****
C.fam FTH1 C18 cDNA ....//GAGTATCTCTTTGACAAGCACACCCTGGGAAACAGTGATAATGAGAGCTAA 552
C.fam FTH1 C11 cDNA ....//GAGTATCTCTTTGACAAGCACACCCTGGGAAACAGTGATAATGAGAGCTAA 552
.......//***
c
5'
FTH1 C11
AATAAAGTAATTTGGTACCCA**(A)**$_{39}$
3'
PolyA Signal
PolyA Tail
d
Possible PolyA Tail
Our Label, NCBI Gene ID
47 bp
A1, N.A. GTTTTTTTCTTCCAGTTCTGCCATAAAATCTAAGTA**AATAAA**TAAATAAATGAATGAATGAATGAATAAATAAATAAATAAATAAAAAA
34 bp
B1,LOC102153989 GTTTTTTTCTTCCAGTTCTGCCATAAATTATA**AATAAA**TAAATAAATAAATAAATAAATAAATAATAAAAAA
55 bp
C1, LOC612257 GTTTTTTTCTTCCAGTTCTGCCTTA**AATAAA**TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAAAA
50 bp
D1, LOC100687930 (N)$_{19}$ GTTTTTTTCTTCCAGTTCTGCCATA**AATAAA**TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA
39 bp
D2, FTH1P18 GTTTTTTTCTTCCAGTTCTGCCATA**AATAAA**TAAATAAATAAATGAATGAATGAATGAATGAATAATAAA
81 bp
D3, LOC612281 GTTTTTTTCTTCCAGTTCTGCCATAAATTATA**AATAAA**TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATATAAATAAATGAATAAATAAATAATAAA
Consensus GTTTTTTTCTTCCAGTTCTGCC*TAAA*****A***AA**AA***A***A***A***A**A***AA

Figure S3. An additional copy of the FTH1 gene is unique to the dog and is a processed pseudogene. (**a**) The two copies of the FTH1 gene present in the dog genome are depicted as black arrows. Exons are indicted as grey arrows, and lines connecting exons indicate introns; FTH1 on chromosome 11 has no introns. (**b**) ClustalW alignment of the CDS of each FTH1 gene showing 100% sequence homology. Only a portion of the alignment is shown. (**c**) FTH1 gene on chromosome 11 is depicted. A polyA signal and tail, hallmarks of mRNA transcripts, are located 3' of the gene. (**d**) ClustalW sequence alignment of dog FTHL loci identified through tBLASTn search. Dotted line indicates the end of each respective locus, as designated by NCBI, except in the case of C1 and D2, where the dotted line indicates the end of exon 2. PolyA signal is bolded, and downstream A-rich region is bracketed. Note tandem TAAA repeat in A-rich region.



	Details	Filter Rationale
Results of Dog FTH1 tBLASTn Query	Genome assemblies were queried with the dog FTH1 peptide sequence in a tBLASTn search. The results of this search make up the 'unfiltered' hits.	N.A.
Filter 1 Combine Nearby Hits	Hits that covered different regions of the query sequence, were in the same orientation, and were separated by less than 250 bp were combined into a single hit.	Allows for small insertions and frameshift mutations.
Filter 2 Match Quality	Hits with an Expect (E) value >9.99E14 were removed.	Selects for hits with high homology to query sequence.
Filter 3 Length Parameters	Hits of lengths less than 450 bp or more than 650 bp were removed.	Selects for intact hits that lack deletions and selects against intron-containing genes.
Filtered tBLASTn Results	The tBLASTn hits remaining after filtering.	N.A.

Figure S4. Schematic of filtering process of tBLASTn hits. The reference genomes of several species were queried with the dog FTH1 peptide sequence using tBLASTn. The returned hits from this search were filtered as indicated, with the goal of selecting for FTHL loci.

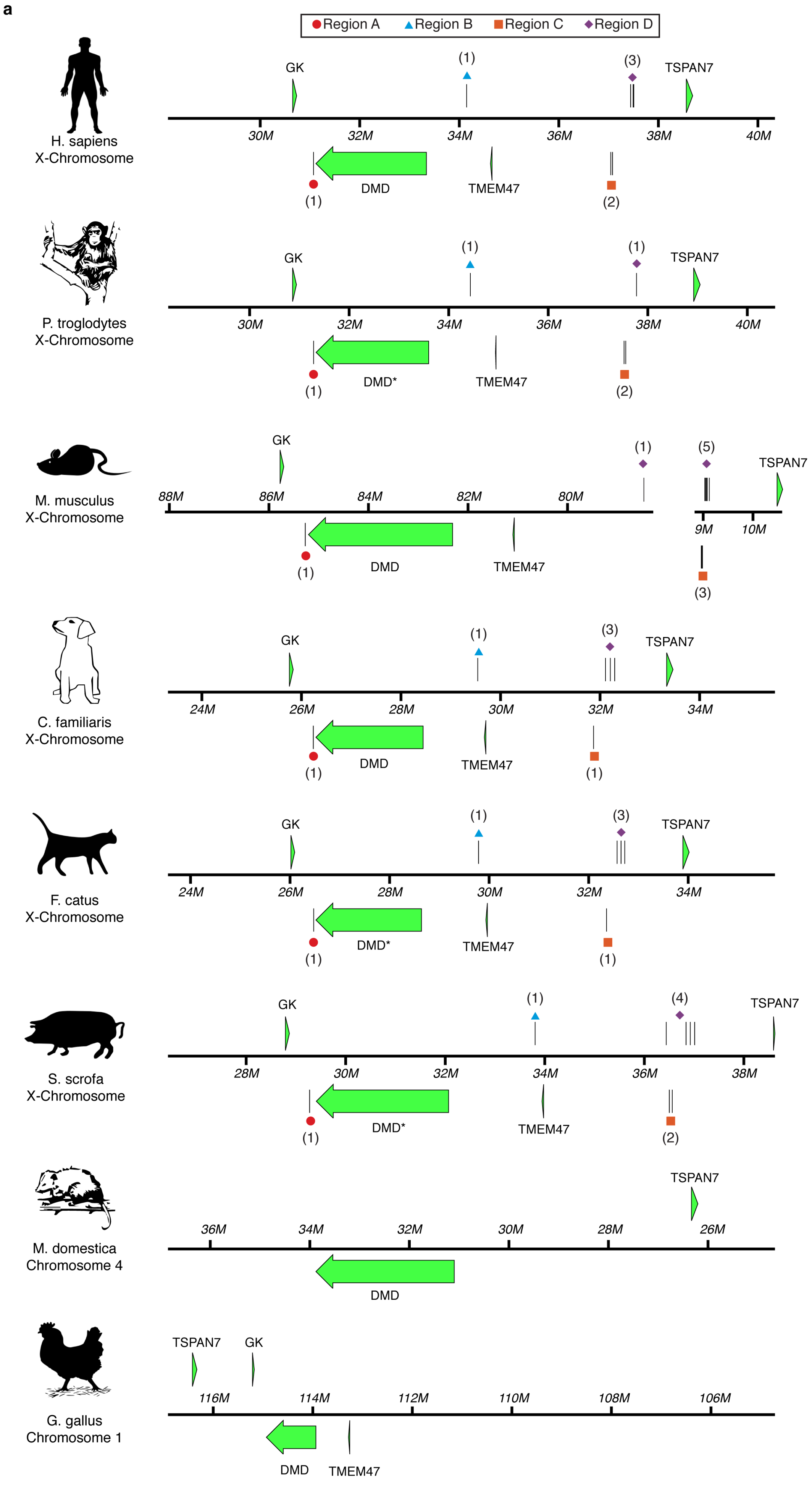


Figure S5. Phylogenetic analysis of identified FTHL loci from tBLASTn search (expanded to additional species). (**a**) Comparison of syntenic portion of dystrophin-containing chromosome from each species showing the grouping of FTHL loci to four regions, labeled A, B, C, and D. The opossum and chicken lack FTHL loci in these regions. Number of pseudogenes present in reach region is indicated in parenthesis.

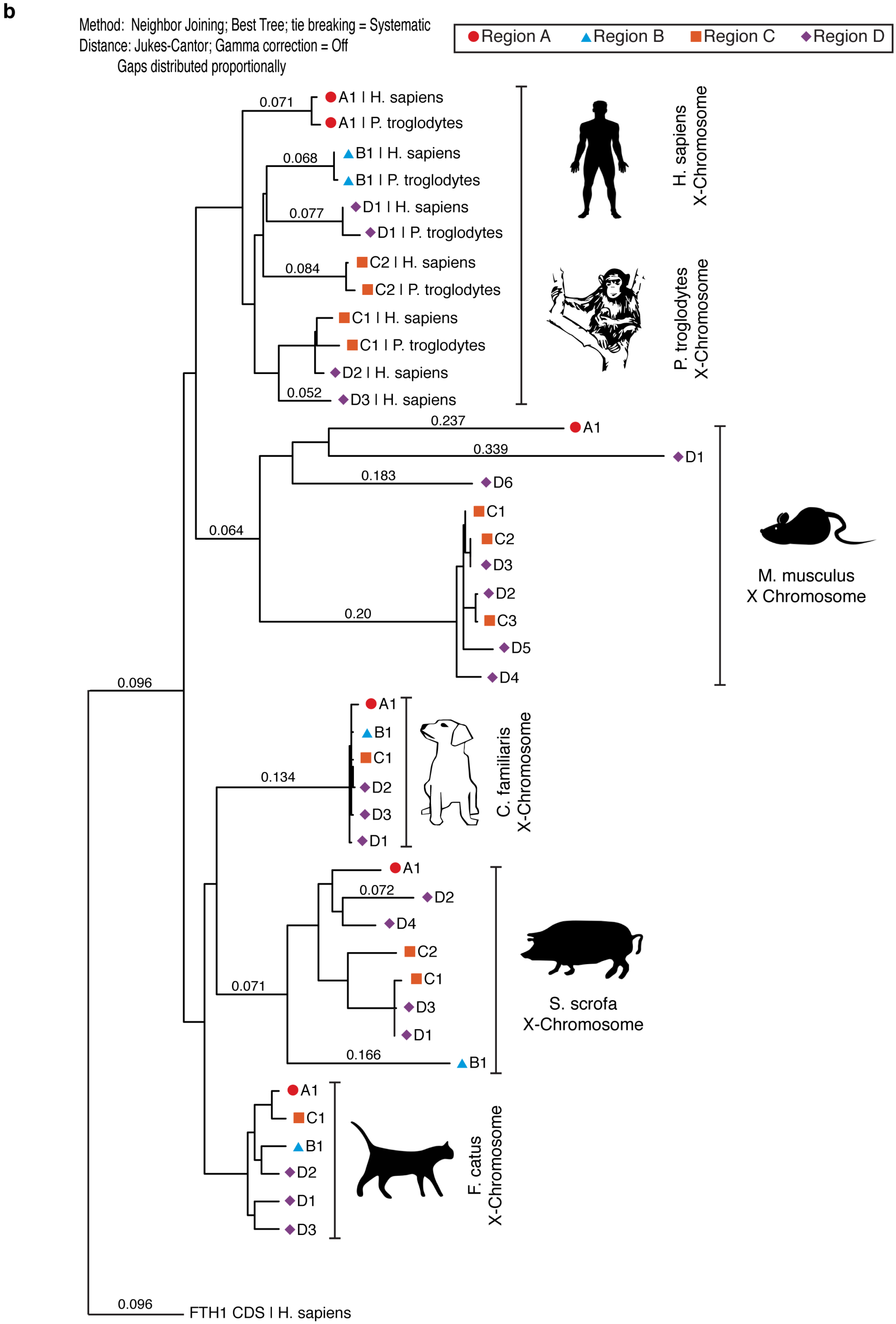


Figure S5. Phylogenetic analysis of identified FTHL loci from tBLASTn search (expanded to additional species). (**b**) Phylogeny of identified FTHL loci. The human FTH1 CDS is used as an outgroup. Branches with values less than 0.05 are not displayed. Tree-building parameters are provided as text in figure. CDS, coding DNA sequence.

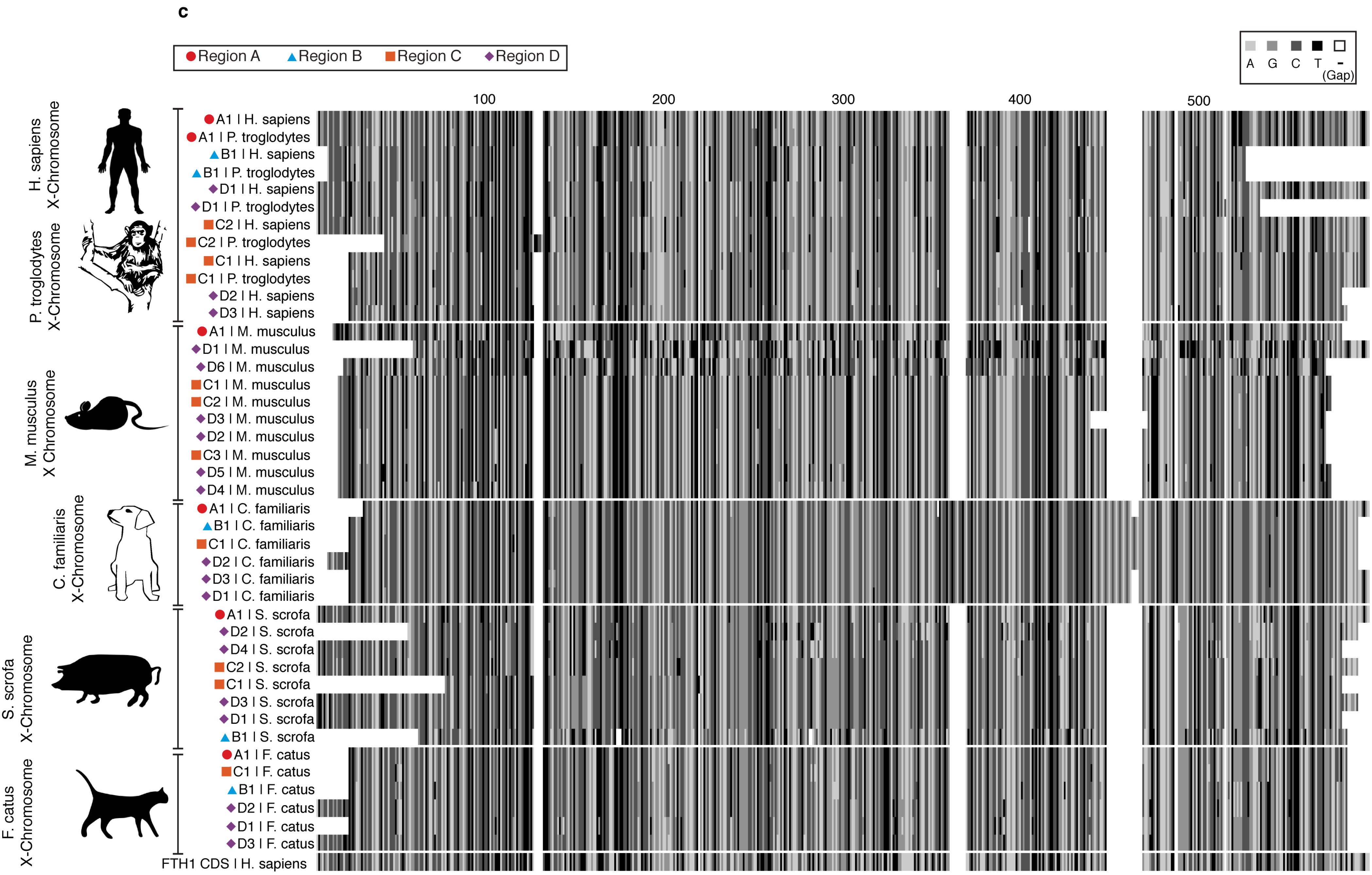


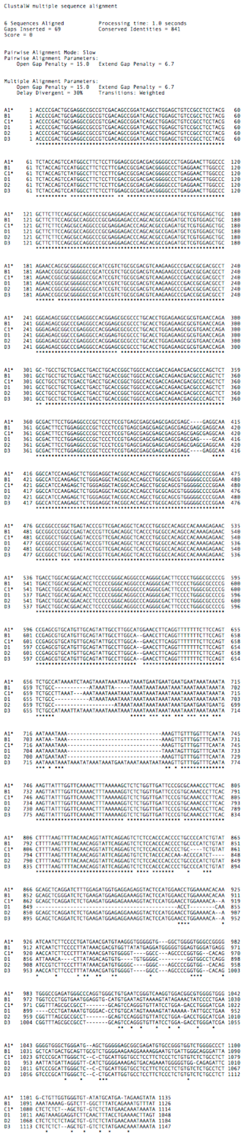
Figure S5. Phylogenetic analysis of identified FTHL loci from tBLASTn search (expanded to additional species). (**c**) ClustalW multiple DNA sequence alignment of identified FTHL loci. Note sequence homology among FTHL loci from all regions is unique to the dog.

```
Sequence: AGap Sequence Range: 1 to 1173

                                10        20        30        40        50        60        70        80        90       100
AGap Sequence           TCCTGAAGGTTCCATGCCAAGGCAATACTGCAACATGCACGCTCGGCGGGGCGCCCAGGGGAAGTCGCCCTGGGCCCTGCCCGGGGGAGGTGTCCGTGCC

                  26238379  26238389  26238399  26238409  26238419  26238429  26238439  26238449  26238459  26238469
CfamX +                 A...................................................................................................>
                         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AGap Sequence           TCCTGAAGGTTCCATGCCAAGGCAATACTGCAACATGCACGCTCGGCGGGGCGCCCAGGGGAAGTCGCCCTGGGCCCTGCCCGGGGGAGGTGTCCGTGCC

                               110       120       130       140       150       160       170       180       190       200
AGap Sequence           AGGTCAGTTCTCTTTGTGGCTGTGGCGCAGGGTGAGCCTGTCGAACGGGTACTCGGCCAGGCCGGCTTCCGGGGCCCCCACGCTGCGCAGGCTGGTGCCG

                  26238479  26238489  26238499  26238509  26238519  26238529  26238539  26238549  26238559  26238569
CfamX +                 ......................................................A...G.........................................>
                        |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||||||||||||
AGap Sequence           AGGTCAGTTCTCTTTGTGGCTGTGGCGCAGGGTGAGCCTGTCGAACGGGTACTCGGCCAGGCCGGCTTCCGGGGCCCCCACGCTGCGCAGGCTGGTGCCG

                               210       220       230       240       250       260       270       280       290       300
AGap Sequence           TAGCCTCCCAGAGCTCTTGGATGGCCTTGCCTCGCTCGCTCGCTCGCTCACGGAGGTAGCGGGCCTCCAGGAAGTCGCAGAGCTGGGCGTCGTTCTGGTC

                  26238579  26238589  26238599  26238609  26238619  26238629
CfamX +                 .................................................G.T>
                        ||||||||||||||||||||||||||||||||||||||||||||||||| |
AGap Sequence           TAGCCTCCCAGAGCTCTTGGATGGCCTTGCCTCGCTCGCTCGCTCGCTCACG
                                                                      26238619  26238629  26238639  26238649  26238659  26238669  26238679
CfamX +                                              G..........................G...........................................>
                                                      |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
AGap Sequence                                        CCTCGCTCGCTCGCTCGCTCACGGAGGTAGCGGGCCTCCAGGAAGTCGCAGAGCTGGGCGTCGTTCTGGTC

                               310       320       330       340       350       360       370       380       390       400
AGap Sequence           GGTGGCCAGCCGGTGCAGGTCAGGTCGAGCAGGCAGGCTCTGGTTCACGCGCTTCTCCAGGTGCAGGGCGCGCTCCGTGGCCCTCGGGCCGCTCTCCCAG

                        26238689  26238699  26238709  26238719  26238729
CfamX +                 .....................................T....T.....>
                        ||||||||||||||||||||||||||||||||||||| |||| |||||
AGap Sequence           GGTGGCCAGCCGGTGCAGGTCAGGTCGAGCAGGCAGGCTCTGGTTCAC

                               410       420       430       440       450       460       470       480       490       500
AGap Sequence           GCGTCGCGGTCGGGCTTCTTGACGTCGCGCAGACGGATGCGGCCCCCGCGCTGGTTCTGCAGCTCCACGAGCATCTCGGCGTGCTGGGTCTCCTCGCGGG

                               510       520       530       540       550       560       570       580       590       600
AGap Sequence           CCTGGCGCTGGAAGAAGCGGGCCAAGTTCCTCAGGGCCCCGTCGTCGCGCTCCAAGGAGAAGGCCATGGACTGGTAGACGTAGGAGGCGGACAGCTCCAG

                               610       620       630       640       650       660       670       680       690       700
AGap Sequence           GCTGATCCGGCTGTCGACGGCGGCCTCGCAGTCGGGGTGCTAGTTCTGGCGAACCTGGGAGATGGGCGCGGCGGCCACGGCTGGCGGCCCGGGCGGCGGG

                               710       720       730       740       750       760       770       780       790       800
AGap Sequence           GCGGGGGCGAGGGCGGGGGCGGGGCGAGGGCGAGGGCGGCCACGGCGCGAGGACAGGCCTGCGGCGCCAATGGCGGTGGCGGCAGGTCTGCGGTTGGTGT

                               810       820       830       840       850       860       870       880       890       900
AGap Sequence           CCAAGCTCGGAGCCCAGGAGAGCCTCGTGGCGTCGCCTGCGGTGCGATGCGGCAGGAGGGAAGTCCCTTAAAGTCCGTTGTTGTGGAGGTGGGAGGTGGA

                               910       920       930       940       950       960       970       980       990      1000
AGap Sequence           ATCCGTTAGGCGGGGGGGCGGGGTCACGACGCAGGGGCGGGGCGAGGTGAGGGGGCGGGGCGTGCTCGGCGGGGGACGGGCGAGGGGGCGTGGCGAGGGA


                      >"Stop"
                          |
                          |   1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
AGap Sequence           GTGGCCGGCGGGGCGGGGCGGGGCACAGGGTGTGGCAGGGCGTGGGGAGGGCGTGGCCGGTGGGCGGGGCGAGGGCGTGGCCGGCGGGCGGGGCGTGGGG

                                                                  26239489  26239499  26239509  26239519  26239529  26239539
CfamX +                                                           .TT.......................................................>
                                                                  |  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
AGap Sequence                                                     TGGGGAGGGCGTGGCCGGTGGGCGGGGCGAGGGCGTGGCCGGCGGGCGGGGCGTGGGG

                              1110      1120      1130      1140      1150      1160      1170
AGap Sequence           CGCGGCGGGGCACCGCTTCAACGTTCCATCTGAGTCTGGGCGGGGCAGGAAGCCAAGGGCAGTGCTGAGTCTC

                          26239549  26239559  26239569  26239579  26239589  26239599  26239609
CfamX +                 .........................................................................>
                        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AGap Sequence           CGCGGCGGGGCACCGCTTCAACGTTCCATCTGAGTCTGGGCGGGGCAGGAAGCCAAGGGCAGTGCTGAGTCTC
```

Figure S6. Sequence alignment of BAC-derived, PCR product and assembly gap region of the dog X-chromosome reference sequence. Alignment corresponds to output from Pustell matrix in Figure S1D.

```
ClustalW multiple sequence alignment

6 Sequences Aligned           Processing time: 0.9 seconds
Gaps Inserted = 4             Conserved Identities = 538
Score = 0

Pairwise Alignment Mode: Slow
Pairwise Alignment Parameters:
    Open Gap Penalty = 15.0   Extend Gap Penalty = 6.7

Multiple Alignment Parameters:
    Open Gap Penalty = 15.0   Extend Gap Penalty = 6.7
    Delay Divergent = 30%     Transitions: Weighted

A1 | C. familia   1                     TCGCCAGAACTAGCACCCCGACTGCGAGGCCGCCGTCGAC  40
B1 | C. familia   1             TCCCAGGTTCGCCAGAACTAGCACCCCGACTGCGAGGCCGCCGTCGAC  48
C1 | C. familia   1             TCCCAGGTTCGCCAGAACTACCACCCCGACTGCGAGGCCGCCGTCGAC  48
D1 | C. familia   1             TCCCAGGTTCGCCAGAACTAGCACCCCGACTGCGAGGCCGCCGTCGAC  48
D2 | C. familia   1 GCGACGCCCATCTCCCAGGTTCGCCAGAACTACCACCCCGACTGCGAGGCCGCCGTCGAC  60
D3 | C. familia   1             TCCCAGGTTCGCCAGAACTACCACCCCGACTGCGAGGCCGCCGTCGAC  48
                                        ************ ***************************

A1 | C. familia  41 AGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACGTCTACCAGTCCATGGCCTTCTCCTTG 100
B1 | C. familia  49 AGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACGTCTACCAGTCCATGGCCTTCTCCTTC 108
C1 | C. familia  49 AGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACGTCTACCAGTCTATGGCCTTCTCCTTC 108
D1 | C. familia  49 AGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACGTCTACCAGTCCATGGCCTTCTCCTTC 108
D2 | C. familia  61 AGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACGTCTACCAGTCCATGGCCTTCTCCTTC 120
D3 | C. familia  49 AGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACGTCTACCAGTCCATGGCCTTCTCCTTG 108
                    ******************************************** **************

A1 | C. familia 101 GAGCGCGACGACGGGGCCCTGAGGAACTTGGCCCGCTTCTTCCAGCGCCAGGCCCGCGAG 160
B1 | C. familia 109 GACCGCGACGACGGGGCCCTGAGGAACTTGGCCCGCTTCTTCCAGCGCCAGGCCCGCGAG 168
C1 | C. familia 109 GACCGCGACGACGGGGCCCTGAGGAACTTGGCCCGCTTCTTCCAGCGCCAGGCCCGCGAG 168
D1 | C. familia 109 GACCGCGACGACGGGGCCCTGAGGAACTTGGCCCGCTTCTTCCAGCGCCAGGCCCGCGAG 168
D2 | C. familia 121 GACCGCGACGACGGGGCCCTGAGGAACTTGGCCCGCTTCTTCCAGCGCCAGGCCCGCGAG 180
D3 | C. familia 109 GAGCGCGACGACGGGGCCCTGAGGAACTTGGCCCGCTTCTTCCAGCGCCAGGCCCGCGAG 168
                    ** *********************************************************

A1 | C. familia 161 GAGACCCAGCACGCCGAGATGCTCGTGGAGCTGCAGAACCAGCGCGGGGGCCGCATCCGT 220
B1 | C. familia 169 GAGACCCAGCACGCCGAGATGCTCGTGGAGCTGCAGAACCGGCGCGGGGGCCGCATCCGT 228
C1 | C. familia 169 GAGACCCAGCACGCCGAGATGCTCGTGGAGCTGCAGAACCGGCGCGGGGGCCGCATCCGT 228
D1 | C. familia 169 GAGACCCAGCACGCCGAGATGCTCGTGGAGCTGCAGAACCGGCGCGGGGGCCGCATCCGT 228
D2 | C. familia 181 GAGACCCAGCACGCCGAGATGCTCGTGGAGCTGCAGAACCGGCGCGGGGGCCGCATCCGT 240
D3 | C. familia 169 GAGACCCAGCACGCCGAGATGCTCGTGGAGCTGCAGAACCGGCGCGGGGGCCGCATCCGT 228
                    **************************************** *******************

A1 | C. familia 221 CTGCGCGACGTCAAGAAGCCCGACCGCGACGCCTGGGAGAGCGGCCCGAGGGCCACGGAG 280
B1 | C. familia 229 CTGCGCGACGTCAAGAAGCCCGACCGCGACGCCTGGGAGAGCGGCCCGAGGGCCACGGAG 288
C1 | C. familia 229 CTGCGCGACGTCAAGAAGCCCGACCGCGACGCCTGGGAGAGCGGCCCGAGGGCCACGGAG 288
D1 | C. familia 229 CTGCGCGACGTCAAGAAGCCCGACCGCGACGCCTGGGAGAGCGGCCCGAGGGCCACGGAG 288
D2 | C. familia 241 CTGCGCGACGTCAAGAAGCCCGACCGCGACGCCTGGGAGAGCGGCCCGAGGGCCACGGAG 300
D3 | C. familia 229 CTGCGCGACGTCAAGAAGCCCGACCGCGACGCCTGGGAGAGCGGCCCGAGGGCCACGGAG 288
                    ************************************************************

A1 | C. familia 281 CGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGAGC-TGCCTGCTCGACCTGACCTGCAC 339
B1 | C. familia 289 CGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGAGCCTGCCTGCTCGACCTGACCTGCAC 348
C1 | C. familia 289 CGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGAGCCTGCCTGCTCGACCTGACCTGCAC 348
D1 | C. familia 289 TGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGAGCCTGCCTGCTCGACCTGACCTGCAC 348
D2 | C. familia 301 CGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGAGCCTGCCTGCTCGACCTGACCTGCAC 360
D3 | C. familia 289 CGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGAGCCTGCCTGCTCGACCTGACCTGCAC 348
                     *********************************** ***********************

A1 | C. familia 340 CGGCTGGCCACCGACCAGAACGACGCCCAGCTCTGCGACTTCCTGGAGGCCCGCTCCCTC 399
B1 | C. familia 349 CGGCTGGCCACCGACCAGAACGACGCCCAGCTCTGCGACTTCCTGGAGGCCCGCTCCCTC 408
C1 | C. familia 349 CGGCTGGCCACCGACCAGAACGACGCCCAGCTCTGCGACTTCCTGGAGGCCCGCTCCCTC 408
D1 | C. familia 349 CGGCTGGCCACCGACCAGAACGACGCCCAGCTCTGCGACTTCCTGGAGGCCCGCTCCCTC 408
D2 | C. familia 361 CGGCTGGCCACCGACCAGAACGACGCCCAGCTCTGCGACTTCCTGGAGGCCCGCTCCCTC 420
D3 | C. familia 349 CGGCTGGCCACCGACCAGAACGACGCCCAGCTCTGCGACTTCCTGGAGGCCCGCTCCCTC 408
                    ************************************************************

A1 | C. familia 400 CGTGAGCGAGCGAGCGAGCGAGCGA----GGCAAGGCCATCCAAGAGCTCTGGGAGGCTA 455
B1 | C. familia 409 CGTGAGCGAGCGAGCGAGCGAGCGAGCGAGGCAAGGCCATCCAAGAGCTCTGGGAGGCTA 468
C1 | C. familia 409 CGTGAGCGAGCGAGCGAGCGAGCGAGCGAGGCAAGGCCATCCAAGAGCTCTGGGAGGCTA 468
D1 | C. familia 409 CGTGAGCGAGCGAGCGAGCGAGCGA----GGCAAGGCCATCCAAGAGCTCTGGGAGGCTA 464
D2 | C. familia 421 CGTGAGCGAGCGAGCGAGCGAGCGAGCGAGGCAAGGCCATCCAAGAGCTCTGGGAGGCTA 480
D3 | C. familia 409 CGTGAGCGAGCGAGCGAGCGAGCGA----GGCAAGGCCATCCAAGAGCTCTGGGAGGCTA 464
                    *************************    *******************************

A1 | C. familia 456 CGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAAGCCGGCCCGGCTGAGTACCCGTTCGA 515
B1 | C. familia 469 CGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAAGCCGGCCCGGCCGAGTACCCGTTCGA 528
C1 | C. familia 469 CGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAAGCCGGCCTGGCCGAGTACCCGTTCGA 528
D1 | C. familia 465 CGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAAGCCGGCCCGGCCGAGTACCCGTTCGA 524
D2 | C. familia 481 CGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAAGCCGGCCTGGCCGAGTACCCGTTCGA 540
D3 | C. familia 465 CGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAAGCCGGCCTGGCCGAGTACCCGTTCGA 524
                    ***************************************** *** **************

A1 | C. familia 516 CAGGCTCACCCTGCGCCACAGCCACAAAGAGAACTGA 552
B1 | C. familia 529 CAGGCTCACCCTGCGCCACAGCCACAAAGAG       559
C1 | C. familia 529 CAGGCTCACCCTGCGCCACAGCCACAAAGAG       559
D1 | C. familia 525 CAGGCTCACCCTGCGCCACAGCCACAAAGAGAACTGA 561
D2 | C. familia 541 CAGGCTCACCCTGCGCCACAGCCACAAAGAG       571
D3 | C. familia 525 CAGGCTCACCCTGCGCCACAGCCACAAAGAGAACTGA 561
                    *******************************
```

Figure S7. ClustalW sequence alignment of FTHL homologous regions present on dog X-chromosome.



```
ClustalW multiple sequence alignment


6 Sequences Aligned          Processing time: 1.0 seconds
Gaps Inserted = 69           Conserved Identities = 841
Score = 0


Pairwise Alignment Mode: Slow
Pairwise Alignment Parameters:
    Open Gap Penalty = 15.0   Extend Gap Penalty = 6.7


Multiple Alignment Parameters:
    Open Gap Penalty = 15.0   Extend Gap Penalty = 6.7
    Delay Divergent = 30%     Transitions: Weighted


A1*       1 ACCCCGACTGCGAGGCCGCCGTCGACAGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACG    60
B1        1 ACCCCGACTGCGAGGCCGCCGTCGACAGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACG    60
C1*       1 ACCCCGACTGCGAGGCCGCCGTCGACAGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACG    60
D1        1 ACCCCGACTGCGAGGCCGCCGTCGACAGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACG    60
D2        1 ACCCCGACTGCGAGGCCGCCGTCGACAGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACG    60
D3        1 ACCCCGACTGCGAGGCCGCCGTCGACAGCCGGATCAGCCTGGAGCTGTCCGCCTCCTACG    60
            ************************************************************


A1*      61 TCTACCAGTCCATGGCCTTCTCCTTGGAGCGCGACGACGGGGCCCTGAGGAACTTGGCCC   120
B1       61 TCTACCAGTCCATGGCCTTCTCCTTCGACCGCGACGACGGGGCCCTGAGGAACTTGGCCC   120
C1*      61 TCTACCAGTCTATGGCCTTCTCCTTCGACCGCGACGACGGGGCCCTGAGGAACTTGGCCC   120
D1       61 TCTACCAGTCCATGGCCTTCTCCTTCGACCGCGACGACGGGGCCCTGAGGAACTTGGCCC   120
D2       61 TCTACCAGTCCATGGCCTTCTCCTTCGACCGCGACGACGGGGCCCTGAGGAACTTGGCCC   120
D3       61 TCTACCAGTCCATGGCCTTCTCCTTGGAGCGCGACGACGGGGCCCTGAGGAACTTGGCCC   120
            ********** ************** ** *******************************


A1*     121 GCTTCTTCCAGCGCCAGGCCCGCGAGGAGACCCAGCACGCCGAGATGCTCGTGGAGCTGC   180
B1      121 GCTTCTTCCAGCGCCAGGCCCGCGAGGAGACCCAGCACGCCGAGATGCTCGTGGAGCTGC   180
C1*     121 GCTTCTTCCAGCGCCAGGCCCGCGAGGAGACCCAGCACGCCGAGATGCTCGTGGAGCTGC   180
D1      121 GCTTCTTCCAGCGCCAGGCCCGCGAGGAGACCCAGCACGCCGAGATGCTCGTGGAGCTGC   180
D2      121 GCTTCTTCCAGCGCCAGGCCCGCGAGGAGACCCAGCACGCCGAGATGCTCGTGGAGCTGC   180
D3      121 GCTTCTTCCAGCGCCAGGCCCGCGAGGAGACCCAGCACGCCGAGATGCTCGTGGAGCTGC   180
            ************************************************************


A1*     181 AGAACCAGCGCGGGGGCCGCATCCGTCTGCGCGACGTCAAGAAGCCCGACCGCGACGCCT   240
B1      181 AGAACCGGCGCGGGGGCCGCATCCGTCTGCGCGACGTCAAGAAGCCCGACCGCGACGCCT   240
C1*     181 AGAACCGGCGCGGGGGCCGCATCCGTCTGCGCGACGTCAAGAAGCCCGACCGCGACGCCT   240
D1      181 AGAACCGGCGCGGGGGCCGCATCCGTCTGCGCGACGTCAAGAAGCCCGACCGCGACGCCT   240
D2      181 AGAACCGGCGCGGGGGCCGCATCCGTCTGCGCGACGTCAAGAAGCCCGACCGCGACGCCT   240
D3      181 AGAACCGGCGCGGGGGCCGCATCCGTCTGCGCGACGTCAAGAAGCCCGACCGCGACGCCT   240
            ****** *****************************************************


A1*     241 GGGAGAGCGGCCCGAGGGCCACGGAGCGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGA   300
B1      241 GGGAGAGCGGCCCGAGGGCCACGGAGCGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGA   300
C1*     241 GGGAGAGCGGCCCGAGGGCCACGGAGCGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGA   300
D1      241 GGGAGAGCGGCCCGAGGGCCACGGAGTGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGA   300
D2      241 GGGAGAGCGGCCCGAGGGCCACGGAGCGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGA   300
D3      241 GGGAGAGCGGCCCGAGGGCCACGGAGCGCGCCCTGCACCTGGAGAAGCGCGTGAACCAGA   300
            ************************** *********************************


A1*     301 GC-TGCCTGCTCGACCTGACCTGCACCGGCTGGCCACCGACCAGAACGACGCCCAGCTCT   359
B1      301 GCCTGCCTGCTCGACCTGACCTGCACCGGCTGGCCACCGACCAGAACGACGCCCAGCTCT   360
C1*     301 GCCTGCCTGCTCGACCTGACCTGCACCGGCTGGCCACCGACCAGAACGACGCCCAGCTCT   360
D1      301 GCCTGCCTGCTCGACCTGACCTGCACCGGCTGGCCACCGACCAGAACGACGCCCAGCTCT   360
D2      301 GCCTGCCTGCTCGACCTGACCTGCACCGGCTGGCCACCGACCAGAACGACGCCCAGCTCT   360
D3      301 GCCTGCCTGCTCGACCTGACCTGCACCGGCTGGCCACCGACCAGAACGACGCCCAGCTCT   360
            ** *********************************************************


A1*     360 GCGACTTCCTGGAGGCCCGCTCCCTCCGTGAGCGAGCGAGCGAGCGAGC----GAGGCAA   415
B1      361 GCGACTTCCTGGAGGCCCGCTCCCTCCGTGAGCGAGCGAGCGAGCGAGCGAGCGAGGCAA   420
C1*     361 GCGACTTCCTGGAGGCCCGCTCCCTCCGTGAGCGAGCGAGCGAGCGAGCGAGCGAGGCAA   420
D1      361 GCGACTTCCTGGAGGCCCGCTCCCTCCGTGAGCGAGCGAGCGAGCGAGCGAG----GCAA   416
D2      361 GCGACTTCCTGGAGGCCCGCTCCCTCCGTGAGCGAGCGAGCGAGCGAGCGAGCGAGGCAA   420
D3      361 GCGACTTCCTGGAGGCCCGCTCCCTCCGTGAGCGAGCGAGCGAGCGAGC----GAGGCAA   416
            *************************************************       ****


A1*     416 GGCCATCCAAGAGCTCTGGGAGGCTACGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAA   475
B1      421 GGCCATCCAAGAGCTCTGGGAGGCTACGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAA   480
C1*     421 GGCCATCCAAGAGCTCTGGGAGGCTACGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAA   480
D1      417 GGCCATCCAAGAGCTCTGGGAGGCTACGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAA   476
D2      421 GGCCATCCAAGAGCTCTGGGAGGCTACGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAA   480
D3      417 GGCCATCCAAGAGCTCTGGGAGGCTACGGCACCAGCCTGCGCAGCGTGGGGGCCCCGGAA   476
            ************************************************************


A1*     476 GCCGGCCCGGCTGAGTACCCGTTCGACAGGCTCACCCTGCGCCACAGCCACAAAGAGAAC   535
B1      481 GCCGGCCCGGCCGAGTACCCGTTCGACAGGCTCACCCTGCGCCACAGCCACAAAGAGAAC   540
C1*     481 GCCGGCCTGGCCGAGTACCCGTTCGACAGGCTCACCCTGCGCCACAGCCACAAAGAGAAC   540
D1      477 GCCGGCCCGGCCGAGTACCCGTTCGACAGGCTCACCCTGCGCCACAGCCACAAAGAGAAC   536
D2      481 GCCGGCCTGGCCGAGTACCCGTTCGACAGGCTCACCCTGCGCCACAGCCACAAAGAGAAC   540
D3      477 GCCGGCCTGGCCGAGTACCCGTTCGACAGGCTCACCCTGCGCCACAGCCACAAAGAGAAC   536
            ******* *** ************************************************


A1*     536 TGACCTGGCACGGACACCTCCCCCGGGCAGGGCCCAGGGCGACTTCCCCTGGGCGCCCCG   595
B1      541 TGACCTGGCACGGACACCTCCCCCGGGCAGGGCCCAGGGCGACTTCCCCTGGGCGCCCCG   600
C1*     541 TGACCTGGCACGGACACCTCCCCCGGGCAGGGCCCAGGGCGACTTCCCCTGGGCGCCCCG   600
D1      537 TGACCTGGCACGGACACCTCCCCCGGGCAGGGCCCAGGGCGACTTCCCCTGGGCGCCCCG   596
D2      541 TGACCTGGCACGGACACCTCCCCCGGGCAGGGCCCAGGGCGACTTCCCCTGGGCGCCCCG   600
D3      537 TGACCTGGCACGGACACCTCCCCCGGGCAGGGCCCAGGGCGACTTCCCCTGGGCGCCCCG   596
            ************************************************************


A1*     596 CCGAGCGTGCATGTTGCAGTATTGCCTTGGCATGGAACCTTCAGGTTTTTTTCTTCCAGT   655
B1      601 CCGAGCGTGCATGTTGCAGTATTGCCTTGGCA--GAACCTTCAGGTTTTTTTCTTCCAGT   658
C1*     601 CCGAGCGTGCATGTTGCAGTATTGCCTTGGCA--GAACCTTCAGGTTTTTTTCTTCCAGT   658
D1      597 CCGAGCGTGCATGTTGCAGTATTGCCTTGGCA--GAACCTTCAGGTTTTTTTCTTCCAGT   654
D2      601 CCGAGCGTGCATGTTGCAGTATTGCCTTGGCA--GAACCTTCAGGTTTTTTTCTTCCAGT   658
D3      597 CCGAGCGTGCATGTTGCAGTATTGCCTTGGCA--GAACCTTCAGGTTTTTTTCTTCCAGT   654
            ********************************  **************************


A1*     656 TCTGCCATAAAATCTAAGTAAATAAATAAATAAATGAATGAATGAATGAATAAATAAATA   715
B1      659 TCTGCC-----------ATAAATTA-----TAAATAAATAAATAAATAAATAAATAAATA   702
C1*     659 TCTGCCTTAAAT---AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA   715
D1      655 TCTGCC-----------ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA   703
D2      659 TCTGCC-------------------ATAAATAAATAAATAAATAAATGAATGAATGAATG   699
D3      655 TCTGCCATAAATTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA   714
            ******                        ***** *** *** *** *** *** *** 


A1*     716 AATAAATAAA------------------------------AAAGTTGTTTGGTTTCAATA   745
B1      703 AATAA-TAAA------------------------------AAAGTTGTTTGGTTTCAATA   731
C1*     716 AATAAATAAA------------------------------AAAGTTGTTTGGTTTCAATA   745
D1      704 AATAAATAAA------------------------------TAAATAGTTTGGTTTCAATA   733
D2      700 AATGAATAAT------------------------------AAAGTTGTTTGGTTTCAATA   729
D3      715 AATAAATAAATAAATATAAATAAATGAATAAATAAATAATAAAGTTGTTTGGTTTCAATA   774
            *** * ***                                ** * **************


A1*     746 AAGTTATTTGGTTCAAAACTTTAAAAAGCTCTCTGGTTGATTCCCGCGCAAACCCTTCAC   805
B1      732 AAGTTATTTGGTTCAAAACTTTAAAAAGGTCTCTGGTTGATTCCCGTGCAAACCCTTCAC   791
C1*     746 AAGTTATTTGGTTCAAAACTTTAAAAAGGTCTCTGGTTGATTCCCGTGCAAACCCTTCAC   805
D1      734 AAGTTATTTGGTTCAAAACTTTAAAAAGGTCTCTGGTTGATTCCCGTGCAAACCCTTCAC   793
D2      730 AAGTTATTTGGTTCAAAACTTTAAAAAGGTCTCTGGTTGATTCCCGTGCAAACCCTTCAC   789
D3      775 AAGTTATTTGGTTCAAAACTTTAAAAAGGTCTCTGGTTGATTCCCGTGCAAACCCTTCAC   834
            *************************** ****************** *************


A1*     806 CTTTTAAGTTTTACAACAGGTATTCAGGAGTCTCTCCACCCACCCCTGCCCCATCTGTAT   865
B1      792 CTTTTAAGTTTTACAACAGGTATTCAGGAGTCTCTCCACCCACCCCTGCCCCATCTGTAT   851
C1*     806 CTTTTAAGTTTTACAACAGGTATTCAGGAGTCTCTCCACCCACCCCTGC----TCTGTAT   861
D1      794 CTTTTAAGTTTTACAACAGGTATTCAGGAGTGTCTCTACCCACCAA-ACCCCATCT----   848
D2      790 CTTTTAAGTTTTACAACAGGTATTCAGGAGTCTCTCCACCCACCCCTGCCCCATCTGTAT   849
D3      835 CTTTTAAGTTTTACAACAGGTATTCAGGAGTCTCTCCACCCACCCCTGCCCCATCTGTAT   894
            ******************************* **** *******    *    ***    


A1*     866 GCAGCTCAGGATCTTTGGAGATGGTGAGGAGAGGTACTCCATGGAACCTGGAAAACACAA   925
B1      852 GCAGCTCGGGATCTCTGAAGATGGAGAGGAAAGGTACTCCATGGAACCTGGAAAACACAA   911
C1*     862 GCAGCTCAGGATCTCTGAAGATGGAGAGAAAAGGTACTCCATGGAACCTGGAAAACA--A   919
D1      849 ---------------------------------------------ACCT--------CAA   855
D2      850 GCAGCTCAGGATCTCTGAAGATGGAGAGGAAAGGTACTCCATGGAACCTGGAAAACA--A   907
D3      895 GCAGCTCAGGATCTCTGAAGATGGAGAGGAAAGGTACTCCATGGAACCTGGAAAACA--A   952
                                                         ****          *


A1*     926 ATCAATCTTCCCCTGATGAACGATGTAAGGGTGGGGGTG---GGCTGGGGTGGGCCGGGG   982
B1      912 ATCAATCTTCCCCTTATAAACGACGTGGTTATATGAGGATGGGGGTGGAGTGGGATGAGG   971
C1*     920 AACCATCTTCCCTTTATAAACGATGTGGGGT----GGCC---AGCCCCGGTGG--CACAG   970
D1      856 ATTAAACA----CTTATAGACAGTGTG----TGTGGGGC---------GGTGGCCTCAGG   898
D2      908 ATCCGTCTTCCCTTTATAAACGATGTGGGGC----GGGC---AGCCCCGGTGG--CACAG   958
D3      953 AACCATCTTCCCTTTATAAACGATGTGGGGT----GGCC---AGCCCCGGTGG--CACAG  1003
            *     *      * **  **   **          *            ****      *


A1*     983 TGGGCCGAGATGGGCCCAGGTGGGCTGTGAATCGGGTCAAGGTGGACGGCGTGGGGTGGG  1042
B1      972 TGGTCCCTGGTGAATGGAGGTG-CATGTGAATAGTAAAAGTATAGAAACTATCCCCTGAA  1030
C1*     971 CGGTTTAGCGCCGCCT-------GCAGTCCAGGGTGTTATCCTGGA-GACCTGGGATCGA  1022
D1      899 ----CCCTGATAAATGTGGGAC-CCTGTGCATAGTAAAAGTATAAAAA-TATTGCCTGAA   952
D2      959 CGGTTTAGCGCCGCCT-------GCAGTCCAGGGTGTTATCCTGGA-GACCTGGCATCGA  1010
D3     1004 CGGTTTAGCGCCGCCT-------GCAGTCCAGGGTGTTATCCTGGA-GACCTGGGATCGA  1055
                                      **  *  *    *   *  *     *    *   


A1*    1043 GGGGTGGGCTGGGATG--AGCTGGGGGAGCGGCGAGATGTGCCGGGTGGTCTGGGGCCCT  1100
B1     1031 GCTCATGACTGCAGTTGCGTCTGGGGAAGAAGGAAAAGGAATGTGATTGGGCAGGGATTA  1090
C1*    1023 GTCCCGCATTGGGCTC--C-CTGCATTGGTGCCTCCTTCTCCCTCTGTGTCTCTGCCTCT  1079
D1      953 AATTATGATTAGGGTT-CATCTGGGGAAAGAGTGACAGAAATGGGGGTGG-CAGAGATTC  1010
D2     1011 GTCCCGCATTGGGCTC--C-CTGCATTGGTGCCTCCTTCTCCCTCTGTGTCTCTGCCTCT  1067
D3     1056 GTCCCGCATTGGGCTC--C-CTGCATTGGTGCCTCCTTCTCCCTCTGTGTCTCTGCCTCT  1112
                     *    *     ***                         *           


A1*    1101 G-CTGTTGGTGGGTGT-ATATGCATGA-TAGAAGTATA  1135
B1     1091 AAATAAAAG-GGTCTT-GGCTTTATCAGAAATGTTTAT  1126
C1*    1080 CTCTCTCT--AGCTGT-GTCTCTATGAACAAATAAATA  1114
D1     1011 AAGTAAAGGAGGTCTTCAACTTTACCTGAAAACTTAGT  1048
D2     1068 CTCTCTCTCTAGCTGT-GTCTCTATGAACAAATAAATA  1104
D3     1113 CTCTCTCT--AGCTGT-GTCTCTATGAACAAATAAATA  1147
               *       *   *    *  *     *        
```

Figure S8. Expanded ClustalW sequence alignment of FTHL homologous regions present on dog X-chromosome. Includes 3' TAAA tandem repeat.

			X-Chromosome Annealing Region		Annealing		PCR Results (Yes/No)	
Primer Pair	Primer Number	Sequence (5'-3')	Start	Stop	Temp. (°C)	Amplicon Size	WT-Male	GSHPMD-Male
1	1	TGGAGTCTCATGGTGCTTTGTTCC	23152290	23152313	61	231	Y	Y
	2	TGCCACAGAGAATTTCACGGAAGG	23152497	23152520				
2	3	TGGGTGAAGCTCGACAACCAGAG	23960538	23960560	62	173	Y	Y
	4	TGGCACCAGGTACAGTTGGGAG	23960689	23960710				
3	5	AAGGCCATAGTGGTCAGGGTGTC	24750336	24750358	63	241	Y	Y
	6	TCAGGGCTCCGATGGGCCTC	24750557	24750576				
4	7	TGCGTGAACGCACACACTAGC	25530610	25530630	62	238	Y	Y
	8	ACACTCCCACACTTGTTGAAGCC	25530825	25530847				
5	9	CCTTCCTGTCTGCCTTGGAG	26154328	26154347	59	403	Y	Y
	10	GTCCATTGGGTGTCCCAGTT	26154711	26154730				
6	11	CTGCACTCATGAAGAGGGGG	26174040	26174059	58	589	Y	Y
	12	GGTAGAGTCTGTGCGTCTGG	26174609	26174628				
7	13	CTCCACTAAAGCACTGCCCA	26184245	26184264	59	355	Y	Y
	14	ACTGGAGCAGGAATCACAGC	26184580	26184599				
8	15	CACTTTCCCCTTTCGCCTCT	26194195	26194214	58	748	Y	Y
	16	GAGAAAACCGAGCTTGCTTCA	26194922	26194942				
9	17	CCTGCTGGCCTATTTGCTGA	26204168	26204187	59	826	Y	Y
	18	ACAGACTCCCCAGAACCTCT	26204974	26204993				
10	19	GCGAGAGCTGGAGTTTGACT	26214222	26214241	58	696	Y	Y
	20	AGGAAGCTGCCTGACAATGTA	26214897	26214917				
11	21	CATCTCCACACAGAAGCCGA	26224113	26224132	59	620	Y	Y
	22	ACTGGTAGCAAGCAAGGCAA	26224713	26224732				
12	23	TGCCTCTTCAAACGCCAGAT	26232627	26232646	59	783	Y	Y
	24	GAGCTTTTCGTGGCTGCAAA	26233390	26233409				
13	25	AAGCCTTGTGCTAGGGAACA	26233493	26233512	58	998	Y	Y
	26	CCTCTTTGTGAGTGTGGAGTGA	26234469	26234490				
14	27	AGCCAATGATGCCACTCTGT	26235187	26235206	59	764	Y	Y
	28	TTCCCCCTCAGAAACGTGTG	26235931	26235950				
15	29	TGAGACCCCGGTACCCATAG	26236154	26236173	59	391	Y	Y
	30	AATGAAGTTTGCCCCGTTGC	26236525	26236544				
16	31	CTGTTCGGCTGAGGTGGATT	26237046	26237065	58	876	Y	Y
	32	CCTGCTGTTGGTGGGTGTAT	26237902	26237921				
17	33	ACCGCTTCAACGTTCCATCT	26239551	26239570	59	295	Y	N
	34	AAAAGGGCCTGAAGGCAGTC	26239826	26239845				
18	35	GTGCTCATCATGATAGGTGTACTC	26241080	26241103	57	897	Y	N
	36	GATACAGCTTATTCTGACCTGGTC	26241953	26241976				
19	37	CAGAGAGGAGCACAAGCAATAA	26242082	26242103	57	601	Y	N
	38	TTGAGTGAATGAGGGACGAAAG	26242661	26242682				
20	39	TCATTCTGTGCATCCCTTCC	26244999	26245018	56	556	Y	N
	40	TTCTGTGGTGCTGTGGTATC	26245535	26245554				
21	41	GAACCCAGTCTTTACTCTTCTCC	26246775	26246797	57	566	Y	N
	42	CTGATAGATATGAGCCACCAATCC	26247317	26247340				
22	43	TAGAATAGCTCCTAGGCAAGAAGG	26250951	26250974	57	908	Y	N
	44	GGGAGACAACTATGATCACAGAAG	26251835	26251858				
23	45	CATAGGGATCCAGACCGATAAATG	26258800	26258823	57	562	Y	N
	46	CAGCTGTACATAGGAGAACATCTC	26259338	26259361				
24	47	GGACTGTCATAGAGCCCTATATAG	26267041	26267064	56	414	Y	N
	48	CTCATCTCTGAGGGAAATACTGAC	26267431	26267454				
25	49	AAAGCCAACTGTGAGCTTGC	26343207	26343226	59	656	Y	N
	50	TCTTCATGACTGCTCCCCCA	26343843	26343862				
26	51	TGGCCTTAGGAAGTGCACAA	26444213	26444232	59	774	Y	N
	52	AAGGGGCAGAATTGGTCGAG	26444967	26444986				
27	53	CCCGTGAACAGGAGTTTGGT	26544472	26544491	59	476	Y	N
	54	TCACCAGGAGAGCTCCTCAA	26544928	26544947				
28	55	GCAATGTGTAAACCTTGTTTCAACT	26644271	26644295	58	685	Y	N
	56	TTGGAGGAGACTTTCCAGCG	26644936	26644955				
29	57	ATCTGCAACCAGTGAACCCT	26744140	26744159	57	819	Y	N
	58	ACCATAGTTTATGCCATGCCT	26744938	26744958				
30	59	CCAATTAGCACTATACACTGCC	27046886	27046907	55	437	Y	N
	60	GGTCCTGGTTTTGCCTATTATC	27047343	27047322				
31	61	CCCATACAAGACTAACTTCCCT	27350933	27350954	55	392	Y	N
	62	CATTGGACTATGTGTAGCGAAG	27351345	27351324				
32	63	CACTTTCAGTCTGTCATCACTG	27650252	27650273	55	866	Y	N
	64	GGCAGATAAATGCTCTGTAGTG	27651138	27651117				
33	65	GTGGAAATCTGCTCTTTAAGGG	27967034	27967055	55	440	Y	N
	66	GAGTGAATCTTCTGAGGTGTTG	27967494	27967473				
34	67	GAATGAGCTAATTGTGGGGATC	28244929	28244950	55	779	Y	N
	68	CGCTAATAGAAAGGAACGTCAG	28245728	28245707				
35	69	GCCCCATAAGAGCAACCCAA	28307181	28307200	59	562	Y	N
	70	ACAGGGCGCTGATAGTCAAA	28307723	28307742				
36	71	TGAAGTGACACTACCTGGGA	28408368	28408387	58	420	Y	N
	72	TTGTAGCTGCTGTGATGGCA	28408768	28408787				
37	73	ACAGGGAGGCAGATACCCTT	28508480	28508499	59	585	Y	N
	74	CAGTGGCAGATGGTGTCGAA	28509045	28509064				
38	75	CACAGCAAGGTTTAGAACCAGT	28612511	28612532	59	516	Y	N
	76	TGTCCTTCCCTTGCTCGTGA	28613007	28613026				
39	77	TGGTTCCTACAACTTCCCCA	28812182	28812201	58	878	Y	N
	78	ATCTCAGTGCACAGGGGTTG	28813040	28813059				
40	79	GGCAGTATGTGCTATGAAGGGA	28914218	28914239	58	513	Y	N
	80	GGATCCTGAGAGCCACTTAGC	28914710	28914730				
41	81	ACCTGGCATAAGAAACTAGCAT	29016674	29016695	57	350	Y	N
	82	CAAGCAAAGGATTTTTGAGAAAGCA	29016999	29017023				
42	83	ATGTTTTCCAGGACAGTTGTGA	29124373	29124394	58	666	Y	N
	84	TGGCTCCCACTCTTTTGAGC	29125019	29125038				
43	85	ATGTGGCTTACTGCTGAGAGG	29224216	29224236	58	835	Y	N
	86	CCTGTGCTGTCCTGATAGCTT	29225030	29225050				
44	87	CGCTACAGTTTGCTGAGTGC	29326541	29326560	57	629	Y	N
	88	AGACAGGTATGTAACTCTCTTCTG	29327146	29327169				
45	89	GTGACAAAGACTCTTCTTGACC	29632191	29632212	55	835	Y	N
	90	GCACTGTCTCCTCTATGGATAA	29633046	29633025				
46	91	GGTCTTTGGTGAGTACTTTTCC	29926906	29926927	55	650	Y	N
	92	CTTCCTGGCAATATGGATGAAG	29927576	29927555				
47	93	AGTCATTAGGTCTTCCAGTCTG	30227121	30227142	56	503	Y	N
	94	AACTCTTAGTCTGAAGTACCGG	30227644	30227623				
48	95	TTCTCTCACCCTAGTCTACTCA	30530437	30530458	56	565	Y	N
	96	GCCTGATATAAAGCACAGGAAG	30531022	30531001				
49	97	CTGTAAAGTGTCTCTGAGTCCT	30826868	30826889	55	482	Y	N
	98	GTACAATCAGGTGCATCAGATC	30827370	30827349				
50	99	TACTTTTCTCAGAGTACCACCC	31134433	31134454	56	617	Y	N
	100	AGAGACCTGGAGTGTCTATAGT	31135070	31135049				
51	101	TGCTTGACAGTTTGGGGAGC	31450263	31450282	59	702	Y	N
	102	CGTTGGAGCCTGATGTCTCA	31450945	31450964				
52	103	GAAGAGGGGACAGCTCTTTCT	31553028	31553048	58	558	Y	N
	104	CCCAACAAGCTCTTTGAGGGA	31553565	31553585				
53	105	CACCTCAGCCGTTTTACTGC	31653058	31653077	58	843	Y	N
	106	GCACAACTGCCATGGAAAGG	31653881	31653900				
54	107	TGCTAGCTGTCTGAGTCCCT	31753413	31753432	59	380	Y	N
	108	TTTGTGTGGCTAATGGGGCT	31753773	31753792				
55	109	CTCCCATCTTGTGAACCTGAGT	31814038	31814059	58	587	Y	N
	110	GCCCTGATAGAGCCAAGAGC	31814605	31814624				
56	111	GGCTGTGTCTATGGCACGTT	31835032	31835051	58	738	Y	N
	112	AGGGGTAGAGGAAATGGTCC	31835750	31835769				
57	113	ACTTGGGATTCCATGGGGGA	31853178	31853197	59	734	Y	N
	114	AGCAAGTTTTCATGGCTGGC	31853892	31853911				
58	115	CCACCCAATAAGCTGGGGAG	31854357	31854376	59	595	Y	N
	116	AACGCATTACCTGATGCCCA	31854932	31854951				
59	117	CCCTTTAAGGCGAGAACCGT	31858072	31858091	57	589	Y	N
	118	AGGAGCTCTCACAGTACAGA	31858641	31858660				
60	119	GCATTTTGAACAAGTACTGGCCT	31859222	31859244	59	558	Y	N
	120	AAAGCCAAAGGCAGTGGTCT	31859760	31859779				
61	121	GGTGGAAACTGCTAGGTGCT	31860531	31860550	58	562	Y	N
	122	CAGAAAGAGAGTGAAATGGGGTT	31861070	31861092				
62	123	TGCTATAAAACAAAGCAGTTGGC	31864701	31864723	58	471	Y	N
	124	TCAAACACGGCTTCCCTGTG	31865152	31865171				
63	125	TCCCATCAGAGAGTGGACCC	31867082	31867101	58	900	Y	N
	126	CCCTGAATGTTAAGCCAGTGA	31867961	31867981				
64	127	CTCTCCCTTCCTCCTGGGTA	31871007	31871026	58	718	Y	Y
	128	AAGGTACGTCAACTAGAGCCA	31871704	31871724				
65	129	CCTTGGGAGATTCCATTGTAGACT	31872250	31872273	58	506	Y	Y
	130	TGTCCTGGAGAACCTTAACACAA	31872733	31872755				
66	131	AACTTTCAGTTCCCATCATCTTTT	31876053	31876076	57	749	Y	Y
	132	ATCAAGCACCTTTTTATGCCAGG	31876779	31876801				
67	133	GGGGAAGGACCACATTGGG	31886019	31886037	59	570	Y	Y
	134	GCTGATTCTTCAAACCATTGGCA	31886566	31886588				
68	135	TCAGACCTCAGAGTATGGGCA	31901007	31901027	58	837	Y	Y
	136	GACCTGTATGCTCCTGAACCT	31901823	31901843				
69	137	TGGGAAGCTGCTCTTCAAAA	31913179	31913198	58	755	Y	Y
	138	GCTCCACACCCAATCTCACA	31913914	31913933				
70	139	GCCCTTTTGACCTCCTCCTC	32013281	32013300	59	458	Y	Y
	140	AGGCCCTCTATGAGGACTGG	32013719	32013738				
71	141	TGCACTTTGCCATTGAGATTCC	32124473	32124494	58	425	Y	Y
	142	TCAATGACAGTTGCTGAAGTTGT	32124875	32124897				
72	143	CCGCTAGAAAGCATCTGGGT	32324343	32324362	59	410	Y	Y
	144	GTGCCCTCTACAGCAAGTGT	32324733	32324752				
73	145	GCAACTACTATGATGAGTTCTAGGC	26237849	26237873	64.5	5632484	N	Y; ~2000bp Amplicon
	146	□CTGGAATTCCTAACCGATTCTCAC	31870309	31870332				
74	147	CCTGAAGGTTCCATGCCAAGGCAAT	26238379	26238403	58	1467	N/A	N/A
	148	AAAAGGGCCTGAAGGCAGTCTGCAC	26239821	26239845				
Sequencing Primers	149	CCTACGTCTACCAGTCCATGGCCTTC						
	150	GACTCAGATGGAACGTTGAAGCG						
	151	CCGCCTAACGGATTCCACCTCC						
	152	GGACTTTAAGGGACTTCCCTCCTG						
	153	GTGAGCCTGTCGAACGGGTACTCAG						
	147	CCTGAAGGTTCCATGCCAAGGCAAT						
	154	GGTGTCCGTGCCAGGTCAGTTCTCT						
	155	TGGGAGAGCGGCCCGAGGGCCACGG						
	146	CTGGAATTCCTAACCGATTCTCAC						
	156	CGAGATGGGCCCAGGTGGGCTGTGA						
	145	GCAACTACTATGATGAGTTCTAGGC						
	157	TTCTGGCGAACCTGGGAGATGGG						
	158	GAAGGCCATGGACTGGTAGACGTAGG						

Figure S9. Detailed primer information. Primer numbers, sequences, annealing locations, annealing temperatures, expected amplicon size, and amplification results are provided.

Data S1. Results from HMMER search of reference proteomes using exon 2 peptide sequence of dog LOC612257 as query. Available as separate .txt file.

Data S2. tBLASTn output for all queried species. Each species is provided in a separate sheet. Detailed information for each hit is provided, and hits remaining after each filtering criteria was applied are indicated in separate columns. Available as separate .xls file.

	Figure		tBLASTn Hit Location			NCBI	NCBI	
Region	Designation	Species	on X-Chromosome		Orientation	Gene Symbol	Gene Type	NCBI Gene Description
A	A1	H. sapiens	31071953	31071402	-	FTHL17	protein coding	ferritin, heavy polypeptide-like 17
	A1	P. troglodytes	31285760	31285209	-	FTHL17	protein coding	ferritin, heavy polypeptide-like 17
	A1	M. musculus	85249677	85270291	+	Fthl17a	protein coding	ferritin, heavy polypeptide-like 17, member A
	A1	C. familiaris	26238702	26238481	-	N.A.: region is within a genome assembly gap		
	A1	F. catus	26476187	26475666	-	FTHL17	protein coding	ferritin, heavy polypeptide-like 17
	A1	S. scrofa	29274687	29275602	-	FTHL17	protein coding	ferritin, heavy polypeptide-like 17
B	B1	H. sapiens	34147040	34147516	+	FTH1P14	pseudo	ferritin, heavy polypeptide 1 pseudogene 14
	B1	P. troglodytes	34435114	34435590	+	LOC473555	pseudo	ferritin heavy polypeptide-like 17
	B1	C. familiaris	29541841	29542399	+	LOC102153989	pseudo	ferritin, heavy polypeptide 1 pseudogene
	B1	F. catus	29786714	29787235	+	LOC101099617	pseudo	ferritin heavy chain pseudogene
	B1	S. scrofa	33806868	33807345	+	LOC100156789	pseudo	uncharacterized LOC100156789
C	C1	H. sapiens	37043556	37043023	-	FTH1P18	protein coding	ferritin, heavy polypeptide 1 pseudogene 18
	C2		37078401	37077851	-	LOC442445	pseudo	ferritin, heavy polypeptide-like 17 pseudogene
	C1	P. troglodytes	37523369	37522836	-	FTH1P18	protein coding	ferritin, heavy polypeptide 1 pseudogene 18
	C2		37558036	37557519	-	LOC473865	pseudo	ferritin heavy polypeptide-like 17
	C1	M. musculus	8962820	8962302	-	Gm5634	protein coding	predicted gene 5634; also known as Fthl17L1
	C2		8976404	8975886	-	Gm14511	protein coding	predicted gene 14511; also known as Fthl17L2
	C3		8986586	8986071	-	Gm14458	protein coding	predicted gene 14458; also known as Fthl17L3
	C1	C. familiaris	31869618	31869060	-	LOC612257	protein coding	ferritin heavy chain-like
	C1	F. catus	32359520	32358999	-	LOC101085694	pseudo	ferritin heavy chain pseudogene
	C1	S. scrofa	36502806	36502341	-	LOC100624935	pseudo	ferritin heavy chain-like
	C2		36562089	36561544	-	LOC100624737	protein coding	ferritin heavy chain-like
D	D1	H. sapiens	37441523	37442074	+	LOC100420326	pseudo	ferritin, heavy polypeptide 1 pseudogene
	D2		37492021	37492539	+	FTH1P19	pseudo	ferritin, heavy polypeptide 1 pseudogene 19
	D3		37505334	37505854	+	FTH1P27	pseudo	ferritin, heavy polypeptide 1 pseudogene 27
	D1	P. troglodytes	37772599	37773090	+	LOC737664	protein coding	ferritin heavy chain
	D1	M. musculus	78470555	78470058	-	Prrg1	protein coding	proline rich Gla (G-carboxyglutamic acid) 1
	D2		9033647	9034162	+	Fthl17	protein coding	ferritin, heavy polypeptide-like 17
	D3		9043736	9044239	+	Gm14499	protein coding	predicted gene 14499; also known as Fthl17L4
	D4		9063176	9063694	+	Gm5635	protein coding	predicted gene 5635; also known as Fthl17L5
	D5		9080053	9080571	+	Gm6826	pseudo	predicted gene 6826; also known as Fthl17L6
	D6		9123467	9123979	+	Gm5753	pseudo	ferritin heavy chain 1 pseudogene
	D1	C. familiaris	32108753	32109313	+	LOC100687930	pseudo	ferritin, heavy polypeptide 1 pseudogene
	D2		32205689	32206259	+	FTH1P18	protein coding	ferritin heavy chain-like
	D3		32293541	32294101	+	LOC612281	pseudo	ferritin, heavy polypeptide 1 pseudogene
	D1	F. catus	32571487	32572008	+	LOC102901113	protein coding	ferritin heavy chain-like
	D2		32650295	32650834	+	LOC101082216	pseudo	ferritin heavy chain pseudogene
	D3		32724236	32724775	+	LOC101082471	pseudo	ferritin heavy chain pseudogene
	D1	S. scrofa	36440389	36440925	+	LOC102167173	protein coding	ferritin heavy chain-like
	D2		36839450	36839944	+	LOC100623926	protein coding	transmembrane gamma-carboxyglutamic acid protein 1-like
	D3		36922815	36923360	+	LOC106506938	protein coding	leucine-rich repeat extensin-like protein 5
	D4		37010756	37011292	+	LOC100625618	protein coding	ferritin heavy chain-like

Table S1. Summary of returned tBLASTn hits from multiple species.