Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Distinct evolutionary affinities of two transcript-supported non- AIM2 paralogues in the canine ALR/PYHIN repertoire

Ryoichi Sugisawa
rsugisawa@med.kindai.ac.jp
Kindai University
Naofumi Tomimoto
Kindai University
Akiyoshi Komuro
Kindai University
Takeshi Ueda
Kindai University
Hitoshi Okada
Kindai University

Research Article

Keywords: canine immunogenomics, PYHIN, IFI16, MNDA, AIM2, comparative genomics

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10746025/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Distinct evolutionary affinities of two transcript-supported non-AIM2 paralogues in the canine ALR/PYHIN repertoire

Ryoichi Sugisawa*, Naofumi Tomimoto, Akiyoshi Komuro, Takeshi Ueda, Hitoshi Okada

Department of Biochemistry, Kindai University Faculty of Medicine, Sakai, Osaka, Japan

*Corresponding author: Ryoichi Sugisawa, Department of Biochemistry, Kindai University Faculty of Medicine, 1-14-1 Mihara-dai, Minami-ku, Sakai, Osaka, 590-0197, Japan.
rsugisawa@med.kindai.ac.jp

ORCID:
Ryoichi Sugisawa, 0000-0002-5912-5647; Naofumi Tomimoto, 0009-0000-5835-2808; Akiyoshi Komuro, 0000-0003-0571-6377; Takeshi Ueda, 0000-0001-9823-6764; Hitoshi Okada, 0000-0001-5597-7212

Acknowledgements

We thank Dr. Daiki Kato (The University of Tokyo) for helpful discussions and advice during the development of this study. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP24K18384 and JP23K19474 and by the 2024 Kindai University Research Enhancement Grant (SR03), all awarded to R.S.

Abstract

The pyrin and HIN domain-containing (PYHIN) family includes proteins involved in inflammasome assembly, DNA-responsive signalling and immune-gene transcription. Rapid lineage-specific diversification, however, makes non-AIM2 gene names unreliable guides to mammalian orthology and function. Dogs were reported to retain two PYHIN genes after AIM2 pseudogenization, but changing annotations left their transcript models and evolutionary affinities unresolved. Here we integrated five canine assemblies, peripheral blood mononuclear cell RNA-sequencing, conserved-domain analysis and domain-resolved phylogenies. We resolved two transcriptionally supported loci encoding complete non-AIM2 PYHIN proteins. RNA-sequencing supported a 468-amino-acid Candidate A representative, an extended Candidate B 5′ structure and an assembly-concordant AN-type 441-amino-acid Candidate B protein. Candidate C lacked a PYD, contained only a partial HIN-like match and lacked uniquely mapped support for its annotated splice chain, providing no evidence of an intact canine AIM2 locus. The expanded HIN phylogeny placed the single Candidate A HIN domain within a carnivoran IFI16-labelled/HIN2-related group and Candidate B within a Canidae MNDA-associated group; both focal placements persisted across alternative substitution models and the gap-filtered analysis, but neither established equivalence to a human paralogue. These sequence-defined proteins provide a compact comparative framework for distinguishing domain-level evolutionary affinity from whole-protein orthology in a rapidly diversifying mammalian gene family.

Keywords: canine immunogenomics; PYHIN; IFI16; MNDA; AIM2; comparative genomics

Introduction

The pyrin and HIN domain-containing (PYHIN) protein family, historically referred to as the absent in melanoma 2 (AIM2)-like receptors (ALRs), comprises a rapidly evolving group of proteins that typically contain an N-terminal pyrin domain (PYD) and one or two C-terminal HIN-200 domains (Brunette et al. 2012; Cridland et al. 2012). The HIN-200 designation originated from the haematopoietic expression, interferon inducibility and nuclear localisation of the initially described proteins, together with their characteristic approximately 200-amino-acid domain (Ludlow et al. 2005). The receptor-based ALR concept was established principally through AIM2, which couples cytosolic double-stranded DNA recognition to inflammasome activation through apoptosis-associated speck-like protein containing a CARD (ASC) (Feriotti et al. 2022; Fernandes-Alnemri et al. 2009; Hornung et al. 2009; Rathinam et al. 2010), and IFI16, which was proposed to link intracellular DNA recognition to stimulator of interferon genes (STING)-dependent type I interferon signalling (Almine et al. 2017; Dunphy et al. 2018; Gamble et al. 2026; Gray et al. 2016; Unterholzner et al. 2010). Subsequent studies, however, have shown that the family cannot be understood solely as a group of DNA receptors: IFI16 functions are strongly dependent on cellular and subcellular context, whereas other members, including MNDA and PYHIN1, regulate immune-gene transcription and antiviral restriction without necessarily acting as direct DNA sensors (Gu et al. 2022; Massa et al. 2020). Thus, PYHIN provides the more neutral family-level designation, whereas ALR denotes the historical receptor-centred framework.

The human ALR/PYHIN locus encodes four HIN-containing PYHIN proteins—AIM2, IFI16, MNDA and PYHIN1—as well as the PYD-only protein POP3 (Connolly and Bowie 2014). In contrast, the C57BL/6J mouse ALR/PYHIN locus contains 13 genes (Brunette et al. 2012). Comparative genomic analyses have further revealed extensive lineage-specific duplication, loss and domain rearrangement across mammals, with AIM2 representing the only clear one-

to-one orthologue shared between the human and mouse repertoires (Brunette et al. 2012; Cridland et al. 2012). Consistent with this evolutionary divergence, Baran et al. showed that the lineage-specific murine PYHIN protein IFI207 regulates cytokine transcription through IRF7 without acting as a DNA sensor (Baran et al. 2023). Thus, the functions of non-AIM2 PYHIN proteins cannot be inferred reliably from human gene names or overall domain architecture alone.

The dog provides a particularly informative system in which to address this problem, both as a major companion-animal species for which accurate interpretation of innate immune pathways has direct veterinary relevance and as a phylogenetically distinct mammal for comparative analysis of PYHIN protein function. The canine lineage retains only two intact ALR/PYHIN genes (Brunette et al. 2012; Cridland et al. 2012) and lacks functional AIM2 following a pseudogenization event that predates the dog-cat divergence, providing a natural contrast to the expanded murine and differently organised human loci (Brunette et al. 2012). The two retained genes were described using study-specific labels—Dog1 and Dog2 by Cridland et al. and gene 1 and gene 2 by Brunette et al.—and neither study established simple one-to-one orthology with human IFI16, MNDA or PYHIN1 (Brunette et al. 2012; Cridland et al. 2012). This ambiguity persists in current genome annotations, in which IFI16- and MNDA-associated names are inconsistently distributed across the two neighbouring loci and their transcript models (O'Leary et al. 2016). Subsequent work has documented responses of canine cells to intracellular DNA stimuli, characterised canine STING signalling and placed canine inflammasome activity within the broader evolutionary remodelling of inflammatory cell-death pathways in Carnivora (Cui and Zhang 2021; Digby et al. 2021; Luff et al. 2013, 2014; Zhang et al. 2017). Within this literature, putative canine PYHIN molecules have been discussed using annotation-derived designations such as canine IFI16 or IFI16/p203, without establishing one-to-one orthology with the corresponding human protein (Cui and Zhang

2021; Luff et al. 2013, 2014). More recently, canine adenovirus type 1 (CAV-1)-induced PANoptosis in canine splenocytes was attributed to AIM2 inflammasome activation (Li et al. 2026), despite genomic evidence that the canine AIM2 locus is pseudogenic (Brunette et al. 2012; Cridland et al. 2012). Thus, the central unresolved issue is not whether canine cells respond to intracellular DNA, but whether—and, if so, how—any component of those responses can be attributed to the evolutionarily defined canine PYHIN repertoire. A current reconstruction integrating genome annotation, transcript evidence and domain-resolved phylogeny is therefore required before the retained canine proteins can be interpreted mechanistically or selected for cross-species functional analysis.

To resolve these uncertainties, we reassessed the canine ALR/PYHIN repertoire by integrating current genome annotations with publicly available RNA-sequencing data, comparative gene-structure analysis, conserved-domain analysis and domain-resolved phylogenetic inference. We examined whether the two historically described canine loci remain supported as distinct transcribed genes, whether an intact AIM2 coding gene is present, and how their predicted protein products relate evolutionarily to mammalian PYHIN lineages independently of current annotation labels. This integrative analysis identified two transcriptionally supported non-AIM2 PYHIN genes, provisionally designated Candidate A and Candidate B, but found no evidence for an intact canine AIM2. Domain-resolved phylogenetic analyses revealed distinct evolutionary affinities for their predicted protein products: the HIN domain of Candidate A showed affinity to the C-terminal HIN domain of IFI16, although Candidate A was not supported as a one-to-one orthologue of human IFI16, whereas Candidate B was placed within a carnivoran MNDA-associated lineage. Previous studies identified two canine PYHIN loci and inferred AIM2 loss, but did not fully reconcile these loci with current annotations, define RNA-sequencing-supported representative coding sequences or resolve their domain-specific evolutionary affinities. Together, these findings

establish a transcript-supported framework for interpreting PYHIN evolution in Canidae and show that annotation-derived product names are insufficient proxies for whole-protein orthology in rapidly diversifying mammalian gene families.

Materials and methods

Reference genomes, annotation records and comparative locus definition

Genome assemblies, gene records, transcript models and protein sequences were obtained from NCBI Gene, RefSeq and NCBI Datasets (Brown et al. 2015; Goldfarb et al. 2025; O'Leary et al. 2024). The domestic dog (*Canis lupus familiaris*; NCBI Taxonomy ID 9615) was the focal species, and the comparative locus schematic additionally included human (*Homo sapiens*) and C57BL/6J mouse (*Mus musculus*) records. UU_Cfam_GSD_1.0 with NCBI *Canis lupus familiaris* Annotation Release 106 (AR106) was used as the primary canine reference. The three focal records were designated Candidate A (LOC488622; GeneID 488622), Candidate B (currently annotated MNDA; GeneID 111089951) and Candidate C (LOC111094466; GeneID 111094466). Comparative records and the five-assembly audit are detailed in the Supplementary Methods and Supplementary Tables S1 and S2.

Conserved-domain analysis and transcriptome analysis

Predicted protein sequences were analysed using NCBI Batch CD-Search against the Conserved Domain Database (CDD) (Marchler-Bauer et al. 2011; Marchler-Bauer and Bryant 2004; Wang et al. 2023). A complete PYHIN coding model was defined as a continuous open reading frame containing complete specific PYD and HIN matches. Public canine peripheral blood mononuclear cell (PBMC) RNA-sequencing data from BioProject PRJNA1035251 (Kaur et al. 2025) were aligned to UU_Cfam_GSD_1.0/AR106 using STAR v2.7.11b (Dobin et al. 2013). Reverse-strand gene counts were summarised as counts per million assigned fragments (CPM), and exact strand-compatible splice junctions were used to assess transcript structures. Run selection, alignment parameters and transcript-specific procedures are provided in the Supplementary Methods and Supplementary Table S3.

Representative canine protein reconstruction

Candidate A models were compared across complete annotated intron chains using uniquely mapped split reads; selection criteria and the coordinate key are provided in the Supplementary Methods, Supplementary Fig. S1 and Supplementary Table S4. Candidate B was reconstructed from cross-assembly transcript models and fragment-level evidence at the local AR106 coding discrepancy. Only alignments with an NH tag of 1 and MAPQ 60 were retained after excluding unmapped, secondary, QC-failure, duplicate and supplementary alignments. Upstream transcript structure and local coding evidence were evaluated independently; reconstruction and translation procedures are detailed in the Supplementary Methods.

Domain-resolved phylogenetic analysis

Complete CDD-bounded PYD and HIN sequences were aligned separately with MAFFT v7.526 using L-INS-i (Katoh and Standley 2013). The expanded datasets comprised 25 PYD sequences (74 sites) and 29 HIN sequences (174 sites). Maximum-likelihood trees were inferred with IQ-TREE v3.1.3 (Wong et al. 2026) using the ModelFinder-selected JTT+G4 (PYD) and Q.PLANT+F+G4 (HIN) models (Kalyaanamoorthy et al. 2017; Minh et al. 2021). Branch support was assessed with 1,000 SH-like approximate likelihood-ratio test (SH-aLRT) replicates and 1,000 ultrafast bootstrap (UFBoot) replicates, with bootstrap nearest-neighbour interchange (BNNI) optimisation applied to UFBoot (Guindon et al. 2010; Hoang et al. 2018). Sensitivity analyses tested alternative empirical substitution models and removal of columns with less than 50% occupancy. Recovery of the prespecified Candidate A and Candidate B focal groups was compared across analyses. Accession-level metadata, exact sensitivity models and full analysis parameters are provided in the Supplementary Methods and Supplementary Table S5.

Data summarisation and visualisation

Expression, splice-junction and fragment-classification results were summarised descriptively. No inferential statistical tests or multiple-testing corrections were performed because each breed-condition combination comprised one pooled unit. Data processing and visualisation used R v4.5.2 and Python v3.12.13. OpenAI ChatGPT and Codex (OpenAI, San Francisco, CA, USA) assisted with scripting and English-language editing. All AI-assisted outputs were reviewed and verified by the authors.

Results

Comparative genomic reconstruction supports two complete canine PYHIN coding loci

To establish the genomic framework for reassessing the canine ALR/PYHIN repertoire independently of current annotation labels, we first examined the conserved *CADM3–SPTA1* interval. The organisation of this interval differs markedly among human, C57BL/6J mouse and dog. The human interval encodes four HIN-containing PYHIN proteins and the PYD-only protein PYDC5/POP3, whereas the mouse interval contains 13 protein-coding PYHIN genes. The corresponding dog interval contains three focal annotation records, termed Candidates A, B and C (Fig. 1a).

In AR106, Candidate A (LOC488622; NC_049259.1:23,728,891–23,768,811) and Candidate B (currently annotated *MNDA*; 23,776,349–23,783,901) were separate adjacent minus-strand loci. Candidate C (LOC111094466; 23,711,793–23,738,083) was annotated on the plus strand and overlapped the 3′ portion of Candidate A (Fig. 1b). The multiple Candidate A protein products were alternative models of the same GeneID rather than separate genes. Accession histories connected historical Dog1 to Candidate A and Dog2 to Candidate B; Dog1 and Candidate A had identical PYD and HIN sequences, whereas Dog2 and Candidate B had identical HIN sequences and 97.3% PYD identity (Brunette et al. 2012; Cridland et al. 2012).

Domain analysis distinguished these two loci from Candidate C. Candidate A and Candidate B each contained a continuous open reading frame encoding complete PYD and HIN domains, whereas Candidate C lacked a PYD and contained only a 39-residue nonspecific partial HIN-like match (Fig. 1c). The representative Candidate A and Candidate B architectures shown in Fig. 1c correspond to the models selected through the transcript-level analyses described below. Across five chromosome-level dog assemblies, Candidates A and B consistently retained complete PYD–HIN coding models, whereas no Candidate C model

encoded a continuous protein containing both domains (Supplementary Table S2). Together, the conserved-locus organisation, domain architecture and cross-assembly evidence support two complete canine PYHIN coding loci, whereas Candidate C does not constitute a third complete PYHIN coding locus.

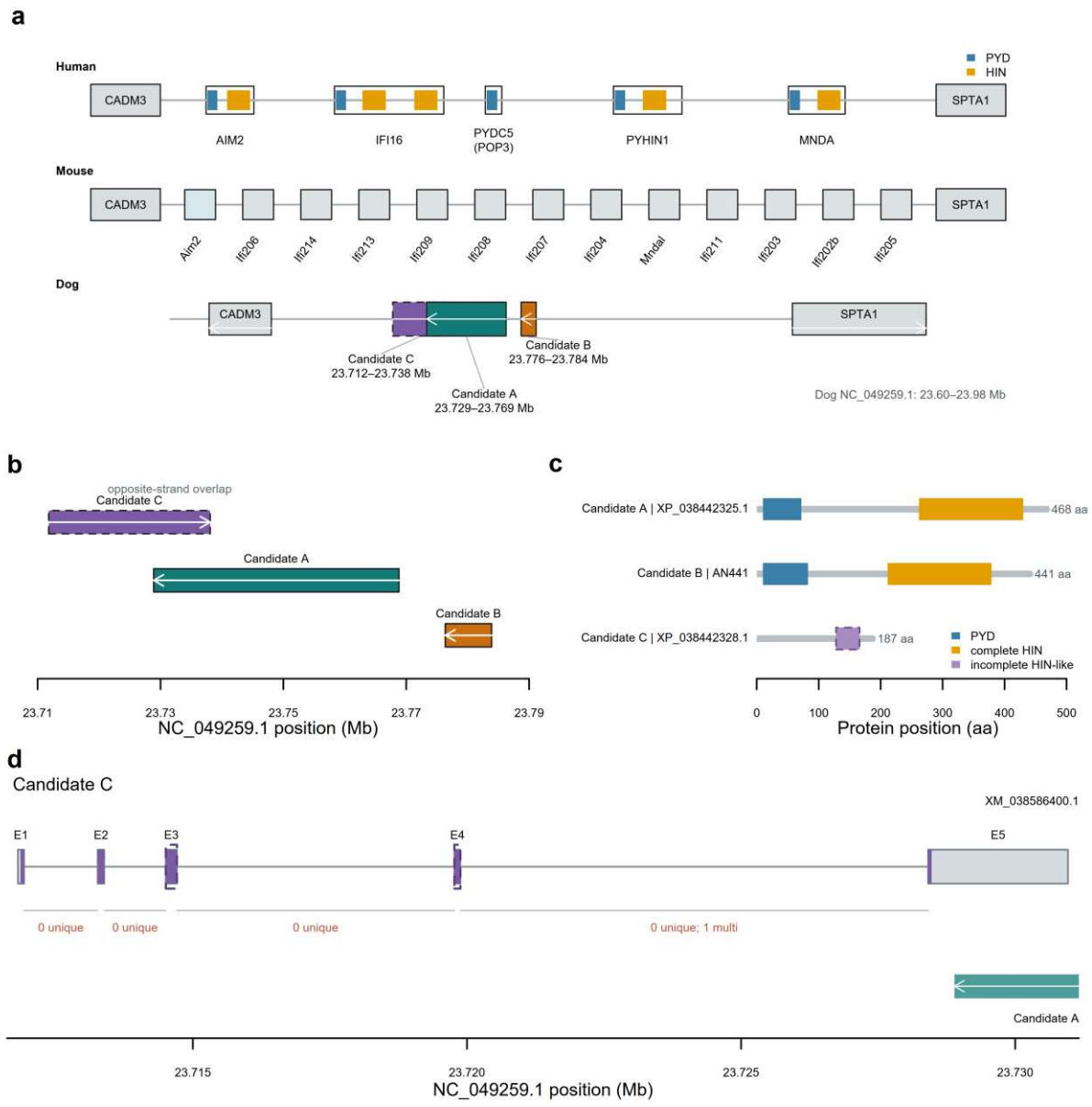


Fig. 1 Comparative genomic and transcriptomic reconstruction of the canine ALR/PYHIN repertoire.

(a) Comparative schematic of the CADM3–SPTA1 interval containing the ALR/PYHIN loci in human, mouse and dog. Individual locus units are enclosed by dark outlines; the dashed Candidate C outline denotes its incomplete PYHIN-like coding model. Human domain architectures are shown for AIM2, IFI16, MNDA, PYHIN1 and the PYD-only protein PYDC5/POP3. Human box widths are approximately proportional to representative protein

length, whereas mouse units remain schematic. The 13 mouse gene names follow current GRCm39 RefSeq records for C57BL/6J mouse; Ifi203-ps and intervening non-PYHIN loci are omitted. Exact AR106 positions are printed for the three focal dog records; the interspecies schematics are not drawn to a common genomic scale. (b) Enlargement of the current AR106 canine interval on NC_049259.1 using the same unit outlines. Candidate C overlaps the 3′ portion of Candidate A in antisense orientation, whereas Candidate B is a separate adjacent minus-strand record. (c) Representative conserved-domain architectures. Candidate A is shown as XP_038442325.1, Candidate B as the RNA-supported AN441 reconstruction and Candidate C as XP_038442328.1. Candidate A and Candidate B contain complete predicted PYD and HIN domains, whereas Candidate C lacks a PYD and contains only a 39-residue nonspecific partial HIN-like match. The cross-assembly audit is summarised in Supplementary Table S2. (d) Exon structure of the Candidate C protein-coding model XM_038586400.1. Values beneath the four annotated introns are summed STAR split-read counts across the eight PBMC units. No intron had uniquely mapped support, and the terminal intron had only one multimapping split read. Candidate A overlaps the terminal Candidate C exon on the opposite strand. Together with the domain and cross-assembly evidence in panel c and Supplementary Table S2, these data do not support Candidate C as an intact AIM2 coding locus or the annotated Candidate C transcript as an expressed PBMC model. They do not exclude unrelated or noncoding transcription at this interval in other tissues or conditions

PBMC transcript evidence supports two expressed PYHIN loci but not the Candidate C coding model

Because Candidate C was the only additional protein-coding annotation within the canonical canine PYHIN interval, we assessed whether its transcription could support a third, potentially AIM2-related PYHIN model. Candidates A and B were detected in all eight PBMC units. Candidate A ranged from 592 to 708 CPM in Gaddi and from 43 to 137 CPM in Labrador units; Candidate B ranged from 394 to 528 and from 40 to 81 CPM, respectively (Table 1). By contrast, Candidate C-associated counts ranged from 0 to 2.92 CPM. These values were interpreted descriptively because each breed–condition combination represented a single pooled analysis unit. Moreover, the opposite-strand overlap between Candidates A and C meant that gene-level counts alone could not establish expression of the annotated Candidate C transcript.

Table 1. Gene-level abundance of the three focal canine records in public PBMC RNA-sequencing units.

Values are reverse-strand counts per million assigned fragments (CPM). Each breed-condition combination represents one pooled descriptive unit; no inferential comparison was performed. Candidate C overlaps Candidate A on the opposite strand, and its gene-level CPM alone is not transcript-specific evidence.

Breed	Condition	SRA run	Candidate A CPM	Candidate B CPM	Candidate C CPM
Gaddi	Unstimulated	SRR28408367	674.148	527.759	2.921
Gaddi	LPS	SRR28408365	707.747	502.811	0.755
Gaddi	Poly(I:C)	SRR28408366	591.719	416.871	2.780
Gaddi	CpG ODN	SRR28408364	602.081	394.480	2.349
Labrador	Unstimulated	SRR29708343	43.315	39.544	0.000
Labrador	LPS	SRR29708345	98.965	81.234	0.103
Labrador	Poly(I:C)	SRR29708344	137.484	68.339	1.152
Labrador	CpG ODN	SRR29708346	73.507	56.407	0.104

We therefore evaluated the exon chain of the five-exon Candidate C protein-coding transcript XM_038586400.1 using strand-compatible splice junctions. None of its four annotated introns had uniquely mapped split-read support across the eight PBMC units, and only one multimapping split read supported the terminal intron (Fig. 1d). Because Candidate C

overlaps Candidate A on the opposite strand, regional transcription could not be assigned to Candidate C without transcript-specific junction evidence. The PBMC data therefore did not support the annotated Candidate C protein-coding exon chain.

This transcriptomic result was concordant with the structural evidence: Candidate C failed to encode a complete PYD–HIN protein across the five evaluated assemblies (Fig. 1c,d and Supplementary Table S2). Accordingly, the current models of the conserved canine PYHIN interval provided no evidence for an intact canine AIM2 coding locus. These findings do not exclude unrelated noncoding transcription at this interval or transcription of other models in tissues and conditions not examined here, but such transcription would not by itself establish an intact AIM2 coding gene.

Splice-junction evidence supports X5 as the Candidate A representative

Having established Candidate A as an expressed PYHIN locus, we next evaluated which of its six AR106 protein-coding transcript models provided the most conservative representative sequence. X1–X4 differed in their 5′ transcript structures but encoded the same 509-aa protein, whereas X5 and X6 encoded 468- and 412-aa proteins, respectively (Fig. 2a). All three predicted protein classes retained complete PYD and HIN domains, although the HIN domain occupied different positions within the proteins (Fig. 2b). Domain completeness alone therefore did not distinguish a representative Candidate A model.

Transcript-oriented comparison revealed marked differences in support for the model-specific 5′ structures (Fig. 2c). The 257-nt 5′-terminal exon of X1 and the 254-nt 5′-terminal exons of X5 and X6 ended at the same splice boundary and used the same first intron, NC_049259.1:23,742,244–23,746,767, which was supported in all eight PBMC units by 2,765 uniquely mapped split reads. The additional upstream junctions of X2, X3 and X4 received only 45 reads in 5/8 units, 50 reads in 4/8 units and 10 reads in 2/8 units,

respectively (Fig. 2c). The junction immediately downstream of these model-specific extensions remained strongly supported, indicating that the limited evidence applied specifically to the additional upstream structures rather than to the shared transcript region.

Across complete intron chains, the weakest coding-region junction in the shared 509-aa model was detected in 6/8 units with 39 reads, whereas every X5 and X6 intron was detected in all eight units. Their minimum junction supports were 2,765 and 1,572 reads, respectively. The complete audit and coordinate key are provided in Supplementary Fig. S1 and Supplementary Table S4. X5 retained a complete PYD–HIN architecture and had the strongest minimum support among models with all introns detected in every unit; XP_038442325.1 was therefore selected as the conservative Candidate A representative. Junctions supporting alternative X1–X4 structures were also detected, so this selection does not exclude other isoforms, and the short-read junction evidence does not phase a full-length native transcript.

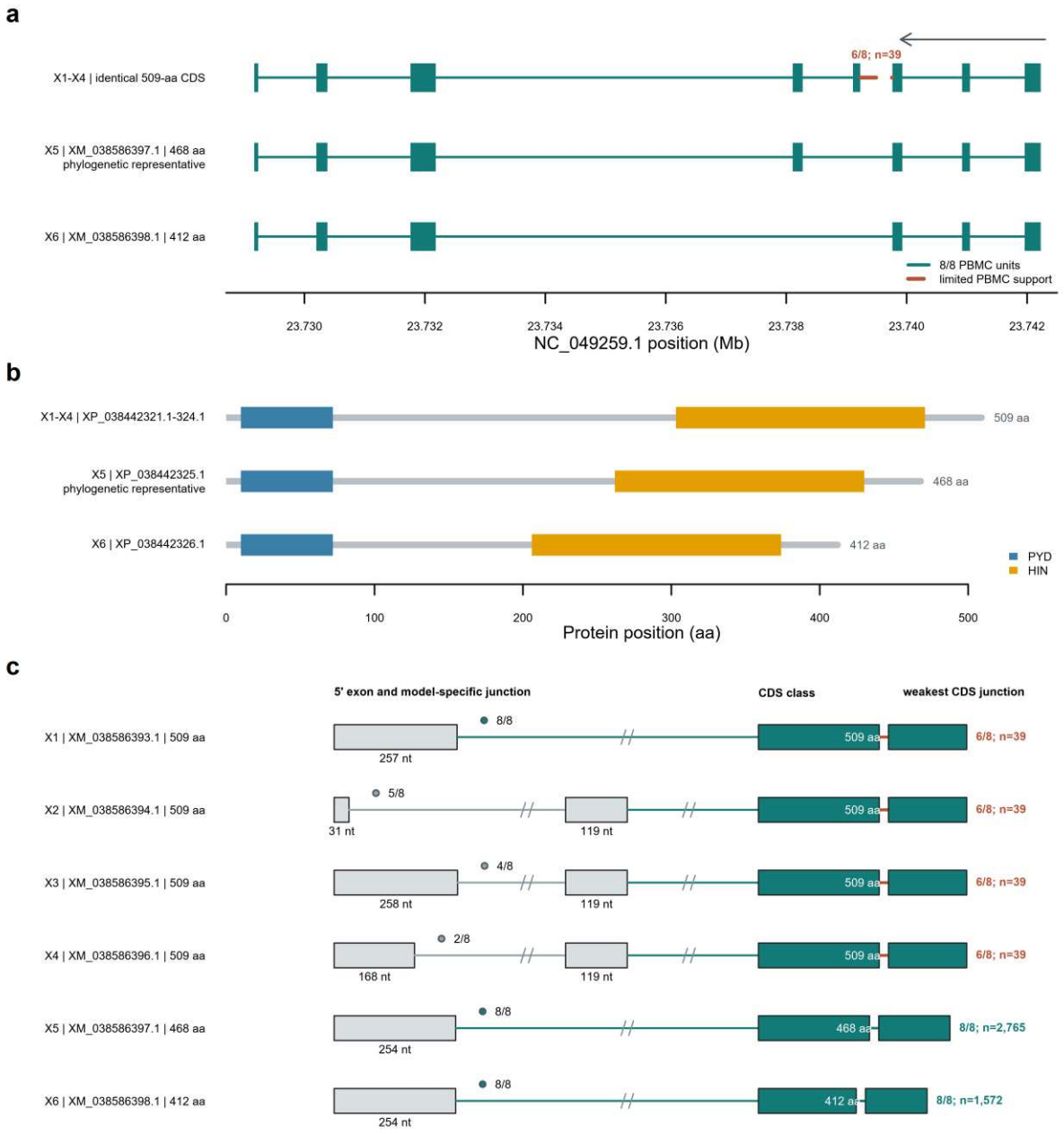


Fig. 2 RNA-seq-supported evaluation of Candidate A transcript and protein models. (a) Coding structures of the three predicted Candidate A protein classes on the AR106 genomic axis. X1–X4 encode the same 509-aa protein. A left-pointing arrow indicates minus-strand transcription. Teal intron connectors denote junctions with uniquely mapped split-read support in all eight PBMC units; red dashed connectors denote junctions with more limited support. The numbers printed above the connectors indicate the number of supporting units and summed uniquely mapped split-read counts. The 468-aa XP_038442325.1 model

selected for phylogenetic inference is identified in the row label. (b) Predicted PYD–HIN architectures of the 509-, 468- and 412-aa protein classes. XP_038442325.1 was selected as the conservative Candidate A representative because it retained a complete PYD–HIN architecture and had the strongest minimum junction support among the models for which every constituent intron was detected in all eight units. Junctions associated with the alternative 509-aa X1–X4 models were nevertheless detected in subsets of PBMC units, and these transcript structures were therefore not excluded. Short-read evidence does not phase the complete transcript on a single native RNA molecule. The full junction-by-model audit and J1–J13 coordinate key are provided in Supplementary Fig. S1 and Supplementary Table S4, respectively. (c) Transcript-oriented comparison of all six Candidate A models. The widths of the 5′-terminal exons reflect relative nucleotide lengths, whereas introns contain break marks and are not drawn to genomic scale. Coloured circles indicate support for the illustrated 5′ junctions. X1 has a 257-nt 5′-terminal exon, whereas X5 and X6 share the same 254-nt 5′-terminal exon. These three exons end at the same splice boundary and use the same first intron (NC_049259.1:23,742,244–23,746,767), which was detected in all eight units with a total of 2,765 uniquely mapped split reads. X2–X4 contain an additional upstream exon and have weaker model-specific junction support, whereas their downstream common 5′ junction remains strongly supported. The connector within each CDS box and the value at right indicate the least-supported coding-region junction for each protein class. The 509-aa X1–X4 class has support in 6/8 units with 39 reads, X5 has support in 8/8 units with 2,765 reads, and X6 has support in 8/8 units with 1,572 reads

RNA-seq and cross-assembly concordance support an AN-type 441-aa Candidate B representative

We next examined Candidate B, for which comparison of public gene models revealed two independent discrepancies involving the 5′ transcript structure and local coding sequence. The public AR106 X2 transcript XM_038586399.1 lacked the extended 5′ structure represented in four non-AR106 dog assemblies (Fig. 3a). Projection of this X1-type structure onto AR106 added a 416-nt untranslated first exon upstream of exon 2; the coding region began 18 nt into exon 2, so the extension altered transcript organisation but not the protein sequence. Its exon 1–exon 2 junction was supported in all eight PBMC units by 1,932 uniquely mapped split reads (Fig. 3b). This established the extended junction but did not exclude a shorter transcript or phase a complete native molecule.

At the local coding discrepancy, the AR106 genomic plus strand contained G at NC_049259.1:23,783,787, corresponding to C in transcript orientation, whereas the public AR106 RNA model excluded this base (Fig. 3c). All 3,087 exact-site-informative uniquely mapped fragments retained this base, and none supported the public-model gap. Among 2,637 fragments spanning the complete window, 1,806 (68.5%) placed a compensating gap at NC_049259.1:23,783,793, 816 (30.9%) at 23,783,794 and 15 (0.6%) elsewhere; the remaining 450 exact-site-informative fragments did not span the complete window and were excluded from these proportions. The two predominant placements generated synonymous threonine codons and thus the same protein sequence; their proportions were not interpreted as allele or isoform frequencies.

At the protein level, the public UU/AR106 sequence XP_038442327.1 encoded CK at residues 37–38, whereas both predominant RNA-supported nucleotide placements and all four non-AR106 assembly proteins encoded AN (Fig. 3d). The resulting AN-type protein contained 441 aa. ROS XP_038304662.1 and Zoey XP_038378526.1 were exact full-length

matches to AN441, whereas the Boxer and UNSW proteins differed from AN441 at one and two positions, respectively; the public UU/AR106 protein differed at two positions. RNA-seq and independent assembly proteins therefore converged on AN441 as the Candidate B representative rather than on the local CK-type sequence distributed with the AR106 transcript model. The X1-type upstream junction and the local AN coding sequence were supported independently but were not phased on a single complete native transcript by the short-read data.

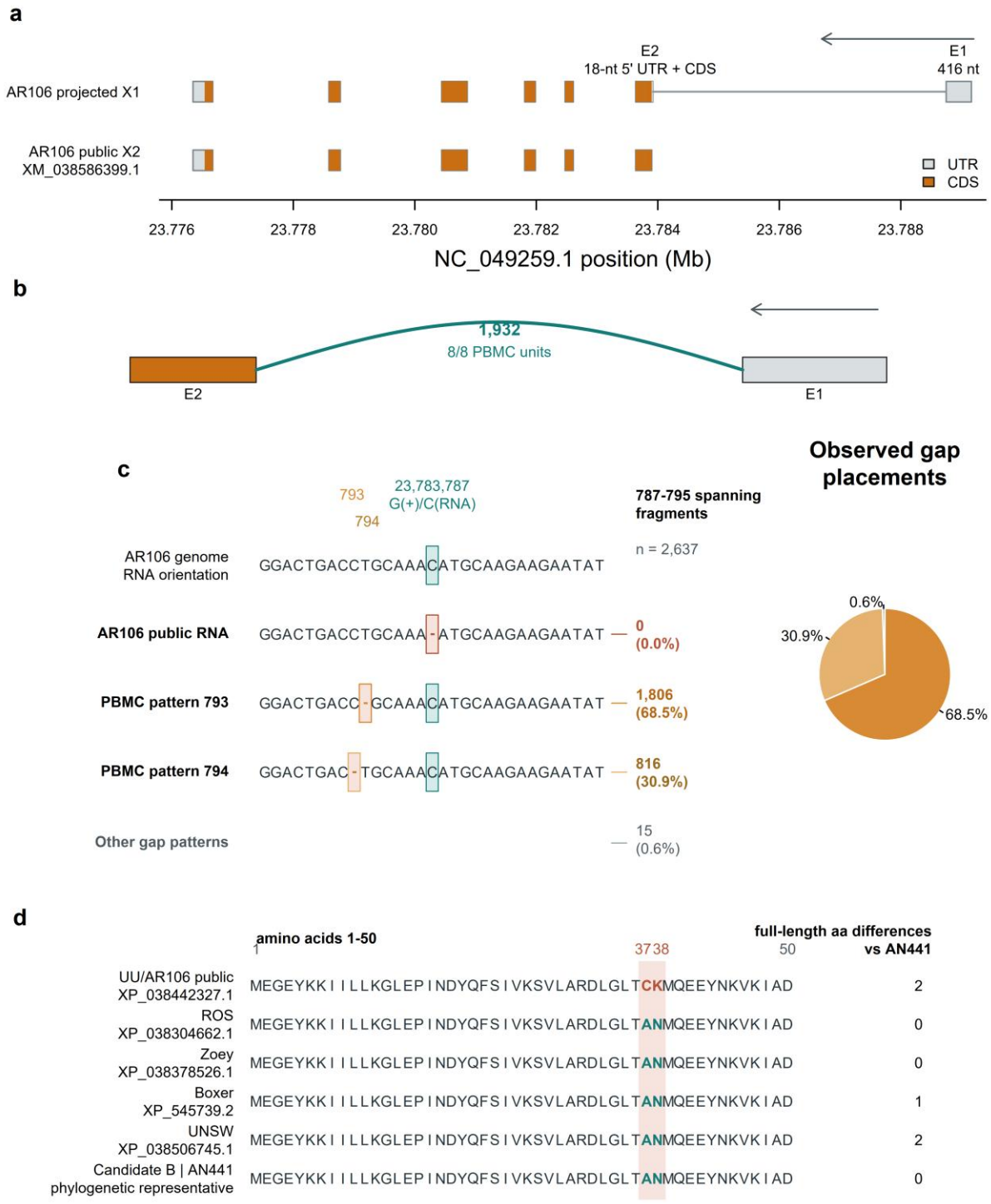


Fig. 3 RNA-seq and cross-assembly evidence support an extended Candidate B transcript structure and an AN-type protein sequence.

(a) AR106 genomic organisation of the public X2 transcript XM_038586399.1 and the projected X1 structure represented in four non-AR106 public dog assemblies. The X1-

specific 416-nt first exon is untranslated, and the coding region begins after an additional 18-nt untranslated segment in exon 2. The arrow indicates minus-strand transcription.

(b) RNA-seq support for the projected X1 exon 1–exon 2 junction. The junction was supported by a total of 1,932 uniquely mapped split reads across all eight PBMC units. This evidence supports the occurrence of the extended X1-type structure but does not exclude a shorter transcript initiating at exon 2.

(c) Integrated transcript-oriented nucleotide alignment and fragment evidence for the independent local coding-model discrepancy. Counts are aligned horizontally with the sequence patterns they support, and the pie chart shows their proportions among the 2,637 complete-window fragments using the same colours. The AR106 genomic plus strand contains G at NC_049259.1:23,783,787, corresponding to C in transcript orientation, whereas the public AR106 RNA model excludes this base. All 3,087 exact-site-informative uniquely mapped fragments retained the focal genomic base, and none supported a gap at the public-model position. Of these fragments, 2,637 spanned the complete NC_049259.1:23,783,787–23,783,795 window: 1,806 (68.5%) placed a one-base gap at 23,783,793, 816 (30.9%) placed it at 23,783,794 and 15 (0.6%) showed another gap pattern. The remaining 450 fragments did not span the complete window and were excluded from these proportions. The two predominant downstream gap placements differ at the nucleotide level but use synonymous threonine codons and therefore produce the same AN-type 441-aa sequence. These proportions describe local alignment patterns and should not be interpreted as allele or isoform frequencies.

(d) Amino-acid alignment of residues 1–50 from the public UU/AR106 protein, four non-AR106 public assembly proteins and the AN441 reconstruction. The public UU/AR106 protein XP_038442327.1 encodes CK at residues 37–38, whereas all four non-AR106 assembly proteins and both predominant RNA-supported nucleotide placements encode AN.

Full-length amino-acid differences from AN441 are shown at right; ROS XP_038304662.1 and Zoey XP_038378526.1 are exact full-length matches. AN441 was selected as the Candidate B sequence for phylogenetic inference. The X1 junction and local AN coding sequence are independently supported by short reads but are not phased across a complete native transcript

Domain-resolved phylogenies support distinct evolutionary affinities for the two canine PYHIN proteins

Having resolved representative sequences for both canine loci, we next examined how their PYD and HIN domains related to established mammalian PYHIN lineages independently of their database labels. Candidate A was represented by the existing X5 protein XP_038442325.1 (468 aa; PYD residues 10–72; HIN residues 262–430), and Candidate B by AN441 (PYD residues 10–83; HIN residues 211–379) (Fig. 4a).

The expanded PYD maximum-likelihood analysis contained 25 sequences and 74 aligned sites (Fig. 4b and Supplementary Fig. S2). Human, mouse and horse AIM2 PYDs formed a separate group that did not include either canine candidate. Candidate A and Candidate B were placed in a four-member Canidae-associated group with red fox and golden jackal MNDA-labelled sequences. This group was recovered across the tested substitution models and after gap filtering, whereas relationships within it were not consistently resolved. Thus, the PYD analysis separated both canine candidates from AIM2 but did not assign their non-AIM2 paralogue identities.

The expanded HIN tree contained 29 sequences and 174 aligned sites and separated the two canine candidates more clearly (Fig. 4c and Supplementary Fig. S3). Candidate B was placed with red fox and golden jackal MNDA-labelled sequences in a Canidae group with 99/100 SH-aLRT/UFBoot support, nested within a broader Caniformia MNDA-associated group with 100/100 support. Candidate A was placed in a five-member carnivoran IFI16-labelled/HIN2-related group containing cat, lion and spotted hyena IFI16-labelled sequences and California sea lion IFI16 HIN2, with 98/98 support. These focal groups persisted across alternative substitution models and the gap-filtered analysis, although deeper HIN topology varied among models. Sequence and domain metadata are provided in Supplementary Table S5. Together, the domain-resolved analyses distinguished the evolutionary affinities of the

predicted protein products encoded by the two transcriptionally supported canine PYHIN loci. Candidate B showed a Canidae MNDA-associated affinity and Candidate A a carnivoran IFI16-HIN2/HIN-B-related affinity, but neither placement established one-to-one orthology with a human PYHIN protein or defined molecular function.

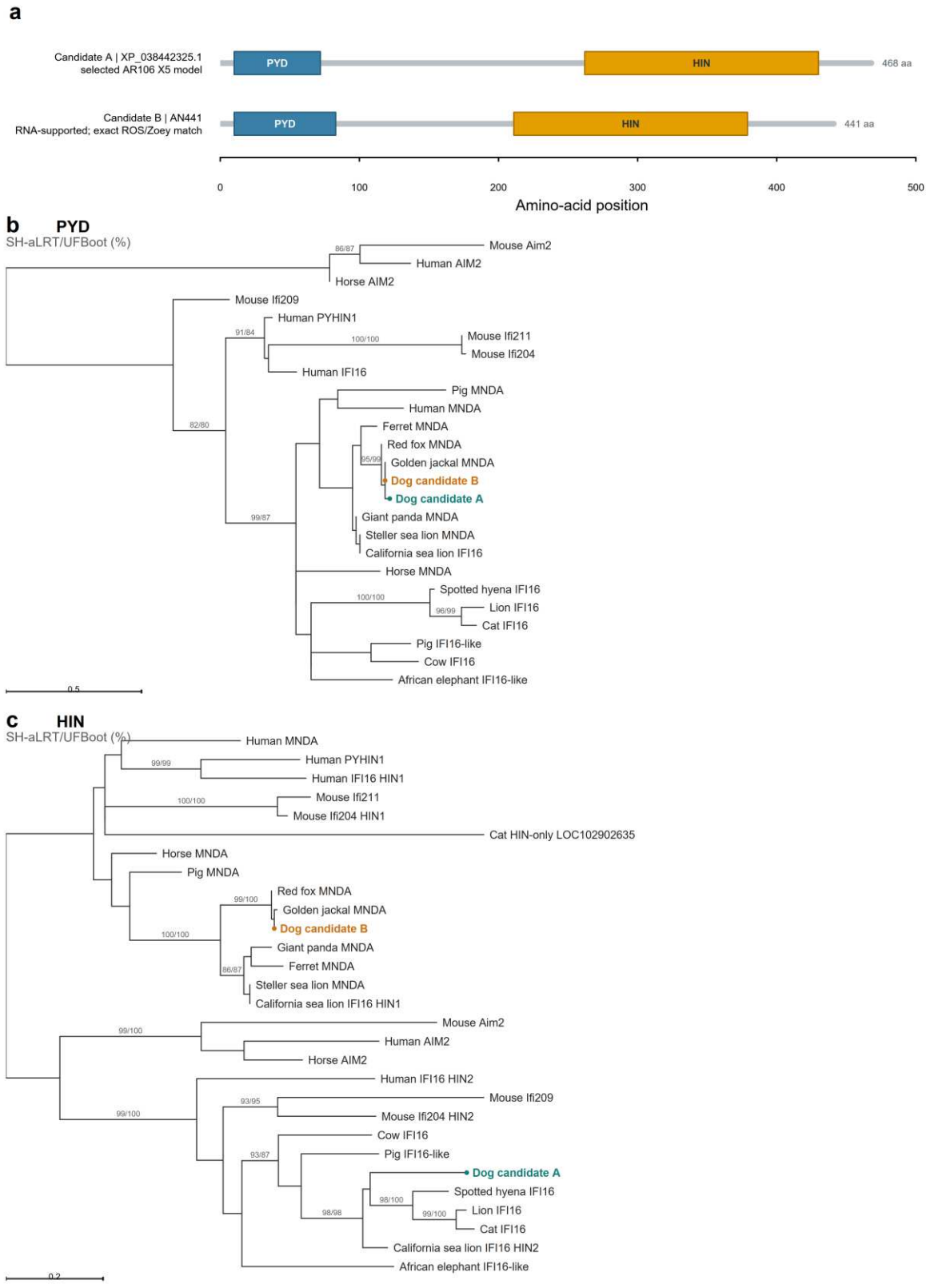


Fig. 4 Representative canine proteins and domain-resolved maximum-likelihood phylogenies of selected mammalian PYHIN proteins.

(a) Representative protein models used for phylogenetic analysis. Candidate A is the existing AR106 X5 protein XP_038442325.1, selected from the six public models based on complete domain architecture and model-wide constituent-junction support (468 aa; PYD, residues 10–72; HIN, residues 262–430). Candidate B is the RNA-supported AN441 reconstruction (441 aa; PYD, residues 10–83; HIN, residues 211–379).

(b) Maximum-likelihood tree inferred from 25 CDD-bounded PYD sequences and 74 aligned amino-acid sites using the ModelFinder-selected JTT+G4 model. Candidate A and Candidate B were placed with red fox and golden jackal MNDA-labelled sequences in a four-member Canidae-associated group, although relationships within this group did not resolve the canine paralogue identities. The very short terminal branches of the canine candidates reflect their inferred near-zero branch lengths and are not clipped graphical elements.

(c) Maximum-likelihood tree inferred from 29 CDD-bounded HIN sequences and 174 aligned sites using the ModelFinder-selected Q.PLANT+F+G4 model. Candidate B was placed with red fox and golden jackal MNDA-labelled sequences in a Canidae group with 99/100 SH-aLRT/UFBoot support, nested within a broader Caniformia MNDA-associated group with 100/100 support. Candidate A was placed with cat, lion and spotted hyena IFI16-labelled sequences and California sea lion IFI16 HIN2 in a carnivoran IFI16-labelled/HIN2-related group with 98/98 support. Coloured terminal points and labels identify the two dog candidates. Branch labels show SH-aLRT/UFBoot percentages from 1,000 replicates of each test; labels are displayed when either support value is at least 80%. Trees are midpoint rooted for display only, and branch lengths denote substitutions per site. Proximity within an individual domain tree was not interpreted by itself as evidence of one-to-one orthology or functional equivalence. The corresponding PYD and HIN alignment overviews are provided in Supplementary Figs. S2 and S3, respectively, and accession-level domain metadata are provided in Supplementary Table S5

Discussion

PYHIN proteins occupy several distinct positions in innate immunity, from AIM2 inflammasome assembly (Fernandes-Alnemri et al. 2009; Hornung et al. 2009; Rathinam et al. 2010) to context-dependent regulation of DNA-responsive signalling (Almine et al. 2017; Dunphy et al. 2018; Gamble et al. 2026; Unterholzner et al. 2010) and immune-gene transcription (Baran et al. 2023; Gu et al. 2022; Massa et al. 2020). Their rapid lineage-specific diversification, however, means that the identities and functions of non-AIM2 paralogues cannot generally be transferred across species by gene name alone (Baran et al. 2023; Brunette et al. 2012; Connolly and Bowie 2014; Cridland et al. 2012; Gray et al. 2016). In dogs, changing annotation labels and uncertain correspondence to human IFI16 and MNDA have complicated molecular interpretation, while the attribution of DNA-responsive phenotypes to AIM2 conflicts with earlier evidence of its pseudogenization. We therefore integrated current genome records, cross-assembly comparisons, PBMC RNA-sequencing evidence and domain-resolved phylogenies to reconstruct the canine PYHIN repertoire independently of database labels. Our analysis supports two expressed canine loci encoding complete non-AIM2 PYHIN proteins, Candidate A and Candidate B, but provides no evidence for an intact AIM2 coding model at Candidate C. This compact two-protein repertoire provides a tractable system for separating domain-level evolutionary relationships from lineage-specific full-length architectures alongside the highly divergent human and mouse repertoires. Candidate A contains a single HIN domain placed within a carnivoran IFI16-labelled/HIN2-related group, whereas the Candidate B HIN domain was placed within a Canidae MNDA-associated group nested in a broader Caniformia group. The persistence of these focal placements across alternative substitution models and the gap-filtered analysis strengthens their use as comparative hypotheses, but they remain domain-level evolutionary affinities rather than evidence of direct equivalence to human IFI16 or MNDA.

This distinction is particularly important for Candidate A, whose IFI16 HIN2/HIN-B-related affinity connects it to a human protein with diverse and incompletely resolved functions. Human IFI16 has been implicated in intracellular DNA recognition and STING-dependent type I interferon signalling (Almine et al. 2017; Unterholzner et al. 2010). Studies of KSHV- and HSV-1-infected cells have also associated nuclear viral-DNA recognition with IFI16-dependent ASC–caspase-1 inflammasome responses (Johnson et al. 2013; Kerur et al. 2011). IFI16 has further been linked to nuclear DNA-damage signalling and replication-fork protection (Dunphy et al. 2018; Gamble et al. 2026), underscoring that its biology is not reducible to a single receptor function. Recent structural studies further qualify a simple inflammasome-receptor model. The IFI16 PYD forms a filament whose architecture differs from those of AIM2, ASC and NLRP3 and appears incompatible with direct heterotypic assembly with ASC PYD filaments, whereas mutations that impair IFI16 filament formation reduce its ability to enhance STING activation (Garg et al. 2025; Takahashi et al. 2026). These findings do not negate the reported IFI16-dependent inflammatory phenotypes, but they leave direct IFI16–ASC nucleation mechanistically unresolved. Domain architecture imposes an additional constraint: human IFI16 contains tandem HIN-A and HIN-B domains with distinct contributions to DNA recognition and interferon induction (Ni et al. 2016), whereas Candidate A contains only one HIN domain related to IFI16 HIN2/HIN-B. Candidate A therefore provides a defined structural comparator for IFI16, but its phylogenetic placement alone does not establish DNA sensing, STING regulation or inflammasome activity.

Candidate B had a clearer placement within a well-supported Canidae MNDA-associated group nested in a broader Caniformia group, although this likewise does not establish functional equivalence to human MNDA. Human MNDA regulates pathogen-induced type I interferon output in monocytes through transcriptional control of IRF7 (Gu et al. 2022),

providing a focused hypothesis for Candidate B rather than an assigned function. The RNA-supported AN441 reconstruction supplies a sequence-defined model for testing that hypothesis: it corrects residues 37–38 within the predicted PYD of the AR106 protein and is identical to the independent ROS and Zoey proteins. The extended 5′ transcript structure and AN441 coding sequence were supported independently, however, and their linkage on a complete native transcript remains to be confirmed. The restraint required for both canine candidates is reinforced by murine PYHIN biology. Although p204 was implicated in intracellular DNA responses (Unterholzner et al. 2010), deletion of the complete murine *Alr* locus did not abolish several tested DNA-induced interferon responses (Gray et al. 2016). Other murine members act at different levels of immune regulation: p205 controls *Asc* transcription (Ghosh et al. 2017), whereas IFI207 has been reported to promote immune-gene transcription (Baran et al. 2023) and, in separate studies, to stabilise STING (Moran et al. 2024) and limit its activation-induced K63-linked ubiquitination and degradation (Enya et al. 2026). Thus, phylogenetic affinity can identify relevant comparisons, but it cannot determine whether a PYHIN protein functions through nucleic-acid sensing, transcriptional control or signal regulation.

The distinction between molecular identity and pathway-level function is particularly relevant to the existing canine DNA-response literature. Canine keratinocytes mount interferon and cytokine responses to transfected poly(dA:dT) (Luff et al. 2013, 2014), and a functional canine STING pathway has been demonstrated (Zhang et al. 2017); however, these pathway-level observations do not identify the upstream PYHIN molecule. Earlier studies described the investigated canine PYHIN transcripts as IFI16 or IFI16/p203 according to the annotations available at the time (Cui and Zhang 2021; Luff et al. 2013, 2014), but those designations do not establish one-to-one orthology or functional equivalence with human IFI16. The same constraint is more consequential for AIM2. Candidate C lacked a complete

PYD–HIN architecture across the evaluated assemblies, and its annotated coding-transcript splice chain was unsupported in the PBMC dataset. These findings are consistent with previous evidence of canine AIM2 pseudogenization (Brunette et al. 2012; Cridland et al. 2012), although they do not exclude noncoding or condition-specific transcription from the interval. The CAV-1-associated cell-death phenotype described as AIM2-dependent PANoptosis (Li et al. 2026) therefore remains a pathway-level observation and does not by itself establish an intact AIM2 protein or its molecular involvement. NLRP3–ASC-associated responses and NLRP3/GSDMD-associated pyroptotic outputs have been reported in other canine infection systems (Tao et al. 2022; Wang et al. 2024), while inflammasome pathways have undergone broader remodelling in Carnivora (Digby et al. 2021). These studies do not identify an alternative cause of the CAV-1 phenotype, but demonstrate that inflammasome-associated outputs in dogs are not unique evidence of AIM2 activity and that other sensing or regulatory routes remain plausible. Assigning the CAV-1 phenotype to any specific molecule will require a sequence-validated target, gene-specific loss-of-function and rescue with a validated coding sequence.

Although the public PBMC short-read data cannot resolve tissue-specific expression or phase full-length transcripts and the molecular functions of the two proteins remain to be experimentally validated, the present analysis provides sequence-defined models for two distinct canine non-AIM2 PYHIN paralogues. Their functions should now be determined on their own molecular terms rather than inferred from human gene names. Functional comparison of these proteins with human and murine PYHIN members can test which activities are associated with conserved PYD and HIN domains and which depend on lineage-specific full-length architectures. Thus, the sequence-defined canine repertoire provides a compact comparative framework for investigating how conserved PYD and HIN domains and lineage-specific full-length architectures have shaped the diversification of

mammalian PYHIN proteins and their immune functions.

Statements and Declarations

Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP24K18384 and JP23K19474 and by the 2024 Kindai University Research Enhancement Grant (SR03), all awarded to R.S.

Competing interests

The authors declare no competing interests.

Ethics approval

Only publicly available genomic and RNA-sequencing data were analysed. No new human or animal experiments were performed, and ethical approval was not applicable.

Author contributions

R.S.: Conceptualization, Methodology, Software, Investigation, Formal analysis, Data curation, Visualization, Funding acquisition, Writing – original draft, Writing – review and editing. N.T., A.K., T.U. and H.O.: Validation, Writing – review and editing. All authors read and approved the final manuscript.

Data availability

All primary sequence and RNA-sequencing data analysed in this study are publicly available from NCBI. The canine genome assemblies used were UU_Cfam_GSD_1.0 (GCF_011100685.1), ROS_Cfam_1.0 (GCF_014441545.1), UMICH_Zoey_3.1 (GCF_005444595.1), UNSW_CanFamBas_1.0 (GCF_013276365.1) and Dog10K_Boxer_Tasha (GCF_000002285.5). The canine PBMC RNA-sequencing data are available under BioProject PRJNA1035251, with individual run accessions listed in Supplementary Table S3. The representative Candidate A protein is XP_038442325.1, and the Candidate B AN441 reconstruction is identical to the public proteins XP_038304662.1 and XP_038378526.1. Accession-level sequence and domain metadata are provided in Supplementary Table S5. Derived cross-assembly, splice-junction and alignment results supporting the conclusions are provided in Supplementary Tables S2, S4–S5 and Supplementary Figs. S1–S3. Processed source tables, domain alignments, tree files and custom scripts have been assembled as a versioned reproducibility package and are available from the corresponding author upon reasonable request. No new sequencing data were generated.

Supplementary Information

Supplementary Methods, Figs. S1–S3 and Tables S1–S5

References

Almine JF, O'Hare CAJ, Dunphy G, et al (2017) IFI16 and cGAS cooperate in the activation of STING during DNA sensing in human keratinocytes. Nat Commun 8:14392. https://doi.org/10.1038/ncomms14392

Baran M, Feriotti C, McGinley A, et al (2023) PYHIN protein IFI207 regulates cytokine transcription and IRF7 and contributes to the establishment of *K. pneumoniae* infection. Cell Reports 42:112341. https://doi.org/10.1016/j.celrep.2023.112341

581 Brown GR, Hem V, Katz KS, et al (2015) Gene: a gene-centered information resource at
582 NCBI. Nucleic Acids Res 43:D36-42. https://doi.org/10.1093/nar/gku1055

583 Brunette RL, Young JM, Whitley DG, et al (2012) Extensive evolutionary and functional
584 diversity among mammalian AIM2-like receptors. J Exp Med 209:1969–1983.
585 https://doi.org/10.1084/jem.20121960

586 Connolly DJ, Bowie AG (2014) The emerging role of human PYHIN proteins in innate
587 immunity: Implications for health and disease. Biochemical Pharmacology 92:405–
588 414. https://doi.org/10.1016/j.bcp.2014.08.031

589 Cridland JA, Curley EZ, Wykes MN, et al (2012) The mammalian PYHIN gene family:
590 Phylogeny, evolution and expression. BMC Evolutionary Biology 12:140.
591 https://doi.org/10.1186/1471-2148-12-140

592 Cui H, Zhang L (2021) Key Components of Inflammasome and Pyroptosis Pathways Are
593 Deficient in Canines and Felines, Possibly Affecting Their Response to SARS-CoV-2
594 Infection. Front Immunol 11:592622. https://doi.org/10.3389/fimmu.2020.592622

595 Digby Z, Tourlomousis P, Rooney J, et al (2021) Evolutionary loss of inflammasomes in the
596 Carnivora and implications for the carriage of zoonotic infections. Cell Reports
597 36:109614. https://doi.org/10.1016/j.celrep.2021.109614

598 Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner.
599 Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

600 Dunphy G, Flannery SM, Almine JF, et al (2018) Non-canonical Activation of the DNA
601 Sensing Adaptor STING by ATM and IFI16 Mediates NF-κB Signaling after Nuclear
602 DNA Damage. Molecular Cell 71:745-760.e5.
603 https://doi.org/10.1016/j.molcel.2018.07.034

604 Enya T, Zhao W, Geetanjali, et al (2026) IFI207 promotes antiviral responses by modulating
605 STING ubiquitination and degradation. mBio 17:e00612-26.
606 https://doi.org/10.1128/mbio.00612-26

607 Feriotti C, Sá-Pessoa J, Calderón-González R, et al (2022) Klebsiella pneumoniae hijacks the
608 Toll-IL-1R protein SARM1 in a type I IFN-dependent manner to antagonize host
609 immunity. Cell Rep 40:111167. https://doi.org/10.1016/j.celrep.2022.111167

610 Fernandes-Alnemri T, Yu J-W, Datta P, et al (2009) AIM2 activates the inflammasome and
611 cell death in response to cytoplasmic DNA. Nature 458:509–513.
612 https://doi.org/10.1038/nature07710

613 Gamble A, Ward TA, Wheeler OPG, et al (2026) IFI16 senses and protects stalled replication
614 forks. Molecular Cell 86:258-272.e6. https://doi.org/10.1016/j.molcel.2025.12.024

615 Garg A, Niedzialkowska E, Zhou JJ, et al (2025) Structural insights into the atypical filament
616 assembly of pyrin domain-containing IFI16. EMBO J 44:7702–7720.
617 https://doi.org/10.1038/s44318-025-00626-7

Ghosh S, Wallerath C, Covarrubias S, et al (2017) The PYHIN Protein p205 Regulates the Inflammasome by Controlling Asc Expression. J Immunol 199:3249–3260. https://doi.org/10.4049/jimmunol.1700823

Goldfarb T, Kodali VK, Pujar S, et al (2025) NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. Nucleic Acids Res 53:D243–D257. https://doi.org/10.1093/nar/gkae1038

Gray EE, Winship D, Snyder JM, et al (2016) The AIM2-like Receptors Are Dispensable for the Interferon Response to Intracellular DNA. Immunity 45:255–266. https://doi.org/10.1016/j.immuni.2016.06.015

Gu L, Casserly D, Brady G, et al (2022) Myeloid cell nuclear differentiation antigen controls the pathogen-stimulated type I interferon cascade in human monocytes by transcriptional regulation of IRF7. Nat Commun 13:14. https://doi.org/10.1038/s41467-021-27701-x

Guindon S, Dufayard J-F, Lefort V, et al (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010

Hoang DT, Chernomor O, von Haeseler A, et al (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281

Hornung V, Ablasser A, Charrel-Dennis M, et al (2009) AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. Nature 458:514–518. https://doi.org/10.1038/nature07725

Johnson KE, Chikoti L, Chandran B (2013) Herpes simplex virus 1 infection induces activation and subsequent inhibition of the IFI16 and NLRP3 inflammasomes. J Virol 87:5005–5018. https://doi.org/10.1128/JVI.00082-13

Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kaur J, Mohan M, Singh B, et al (2025) In-vitro transcriptomic profiling of indigenous Gaddi vis-à-vis exotic Labrador dogs: insights from systems biology. Front Vet Sci 12:1489905. https://doi.org/10.3389/fvets.2025.1489905

Kerur N, Veettil MV, Sharma-Walia N, et al (2011) IFI16 acts as a nuclear pathogen sensor to induce the inflammasome in response to Kaposi sarcoma associated herpesvirus infection. Cell Host Microbe 9:363–375. https://doi.org/10.1016/j.chom.2011.04.008

Li Y, Zhao Y, Wang W, et al (2026) Canine adenovirus type 1 induces PANoptosis in canine splenocytes via AIM2 inflammasome activation. Vet Microbiol 316:110965. https://doi.org/10.1016/j.vetmic.2026.110965

Ludlow LEA, Johnstone RW, Clarke CJP (2005) The HIN-200 family: More than interferon-inducible genes? Experimental Cell Research 308:1–17. https://doi.org/10.1016/j.yexcr.2005.03.032

Luff JA, Yuan H, Kennedy D, et al (2014) Keratinocyte Antiviral Response to Poly(dA:dT) Stimulation and Papillomavirus Infection in a Canine Model of X-Linked Severe Combined Immunodeficiency. PLOS ONE 9:e102033. https://doi.org/10.1371/journal.pone.0102033

Luff JA, Yuan H, Suter MM, et al (2013) Canine keratinocytes upregulate type I interferons and proinflammatory cytokines in response to poly(dA:dT) but not to canine papillomavirus. Vet Immunol Immunopathol 153:177–186. https://doi.org/10.1016/j.vetimm.2013.02.001

Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:W327–W331. https://doi.org/10.1093/nar/gkh454

Marchler-Bauer A, Lu S, Anderson JB, et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–D229. https://doi.org/10.1093/nar/gkq1189

Massa D, Baran M, Bengoechea JA, Bowie AG (2020) PYHIN1 regulates pro-inflammatory cytokine induction rather than innate immune DNA sensing in airway epithelial cells. J Biol Chem 295:4438–4450. https://doi.org/10.1074/jbc.RA119.011400

Minh BQ, Dang CC, Vinh LS, Lanfear R (2021) QMaker: Fast and Accurate Method to Estimate Empirical Models of Protein Evolution. Syst Biol 70:1046–1060. https://doi.org/10.1093/sysbio/syab010

Moran EA, Salas-Briceno K, Zhao W, et al (2024) IFI207, a young and fast-evolving protein, controls retroviral replication via the STING pathway. mBio 15:e01209-24. https://doi.org/10.1128/mbio.01209-24

Ni X, Ru H, Ma F, et al (2016) New insights into the structural basis of DNA recognition by HINa and HINb domains of IFI16. J Mol Cell Biol 8:51–61. https://doi.org/10.1093/jmcb/mjv053

O'Leary NA, Cox E, Holmes JB, et al (2024) Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. Sci Data 11:732. https://doi.org/10.1038/s41597-024-03571-y

O'Leary NA, Wright MW, Brister JR, et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-745. https://doi.org/10.1093/nar/gkv1189

Rathinam VAK, Jiang Z, Waggoner SN, et al (2010) The AIM2 inflammasome is essential for host defense against cytosolic bacteria and DNA viruses. Nat Immunol 11:395–402. https://doi.org/10.1038/ni.1864

Takahashi R, Zhang Z, Fujimura A, et al (2026) The pyrin domain of IFI16 forms a distinct filament structure important for IFI16-mediated enhancement of STING activation. Structure 34:611-619.e4. https://doi.org/10.1016/j.str.2026.01.011

698 Tao P, Ning Z, Zhou P, et al (2022) H3N2 canine influenza virus NS1 protein inhibits canine
699 NLRP3 inflammasome activation. Vet Immunol Immunopathol 252:110483.
700 https://doi.org/10.1016/j.vetimm.2022.110483

701 Unterholzner L, Keating SE, Baran M, et al (2010) IFI16 is an innate immune sensor for
702 intracellular DNA. Nat Immunol 11:997–1004. https://doi.org/10.1038/ni.1932

703 Wang J, Chitsaz F, Derbyshire MK, et al (2023) The conserved domain database in 2023.
704 Nucleic Acids Res 51:D384–D388. https://doi.org/10.1093/nar/gkac1096

705 Wang Z, Guo L, Yuan C, et al (2024) Staphylococcus pseudintermedius induces pyroptosis
706 of canine corneal epithelial cells by activating the ROS–NLRP3 signalling pathway.
707 Virulence 15:2333271. https://doi.org/10.1080/21505594.2024.2333271

708 Wong TKF, Ly-Trong N, Ren H, et al (2026) IQ-TREE 3: phylogenomic inference software
709 using complex evolutionary models. Mol Biol Evol 43:msag117.
710 https://doi.org/10.1093/molbev/msag117

711 Zhang Y, Zhu M, Li G, et al (2017) Identification and function analysis of canine stimulator
712 of interferon gene (STING). Microbial Pathogenesis 113:202–208.
713 https://doi.org/10.1016/j.micpath.2017.10.047

714

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SugisawaRESM1JME.pdf