Supplementary Information

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

* 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

Journal of Molecular Evolution

Contents: Supplementary Methods; Supplementary Figs. S1–S3; Supplementary Tables S1–S5.

Supplementary Methods

This section provides record-level provenance, run-selection criteria and transcript-specific procedures that are summarised in the main-text Materials and methods. References cited in this Supplementary Information are listed in the main manuscript; no separate supplementary reference list is provided.

Reference-record details and comparative locus construction

Database records used for the comparative locus schematic were last verified on 24 July 2026. Canine genomic coordinates were defined on UU_Cfam_GSD_1.0 chromosome record NC_049259.1. Candidate A corresponded to LOC488622 (GeneID 488622; NC_049259.1:23,728,891–23,768,811, minus strand), Candidate B to the record currently annotated *MNDA* (GeneID 111089951; NC_049259.1:23,776,349–23,783,901, minus strand), and Candidate C to LOC111094466 (GeneID 111094466; NC_049259.1:23,711,793–23,738,083, plus strand). The flanking canine records were *CADM3* (GeneID 610435; NC_049259.1:23,619,494–23,650,802, minus strand) and *SPTA1* (GeneID 488616; NC_049259.1:23,912,752–23,979,956, plus strand).

The comparative schematic used human GRCh38.p14 (GCF_000001405.40; NC_000001.11; RS_2025_08), C57BL/6J mouse GRCm39 (GCF_000001635.27; NC_000067.7; RS_2024_02) and dog UU_Cfam_GSD_1.0. The human row comprised *AIM2*, *IFI16*, *PYDC5/POP3*, *PYHIN1* and *MNDA*. Representative human proteins used for domain display were NP_004824.1, NP_001363516.1, NP_001306939.1, NP_689714.2 and NP_002423.1, respectively. The mouse row contained *Aim2*, *Ifi206*, *Ifi214*, *Ifi213*, *Ifi209*, *Ifi208*, *Ifi207*, *Ifi204*, *Mndal*, *Ifi211*, *Ifi203*, *Ifi202b* and *Ifi205* in genomic order. *Ifi203-ps* and other intervening loci were omitted. Human and mouse distances were schematic; the canine

records were positioned proportionally within NC_049259.1:23.60–23.98 Mb. The complete record inventory is provided in Supplementary Table S1.

Historical Dog1 (XP_003434347.1) and Dog2 (XP_545739.2) records were linked to the current canine records using protein-accession histories, GeneID-replacement histories and protein-sequence comparisons.

Cross-assembly gene-model and conserved-domain audit

Candidates A–C were examined independently in UU_Cfam_GSD_1.0, ROS_Cfam_1.0, UMICH_Zoey_3.1, UNSW_CanFamBas_1.0 and Dog10K_Boxer_Tasha. NCBI Datasets gene packages were parsed without merging records across assemblies. GeneID, transcript type, genomic strand and coordinates, coding-sequence coordinates, RefSeq RNA and protein accessions and predicted protein length were retained for each record. Distributed coding sequences were translated with the standard genetic code and checked for an initiation codon, an in-frame terminal codon, internal stops and agreement with the associated RefSeq protein.

AR106 assigned six protein-coding models to Candidate A: XM_038586393.1/XP_038442321.1, XM_038586394.1/XP_038442322.1, XM_038586395.1/XP_038442323.1, XM_038586396.1/XP_038442324.1, XM_038586397.1/XP_038442325.1 and XM_038586398.1/XP_038442326.1. Candidate B was represented by XM_038586399.1/XP_038442327.1 and Candidate C by XM_038586400.1/XP_038442328.1.

Specific NCBI Conserved Domain Database matches to cd08305 and pfam02760 defined PYD and HIN boundaries, respectively (Marchler-Bauer and Bryant 2004; Marchler-Bauer et al. 2011; Wang et al. 2023). A domain was classified as complete only when it was reported as a specific match without an incomplete-boundary flag. Short, nonspecific or boundary-incomplete matches were retained as partial matches but were not counted as complete domains. Candidate C contained a 39-residue nonspecific pfam02760 match and lacked a

PYD. The assembly-level audit is reported in Supplementary Table S2, and the CDD intervals used for phylogenetic analysis are listed in Supplementary Table S5.

PBMC run selection and alignment parameters

The eight retained pooled PBMC units represented unstimulated, lipopolysaccharide, poly(I:C) and CpG oligodeoxynucleotide conditions for Gaddi and Labrador dogs (Kaur et al. 2025). The corresponding runs, BioSamples and experiments are listed in Supplementary Table S3. When two Labrador runs were available for the same BioSample, the leading 200,000 read pairs from each run were aligned under identical conditions. The four earlier SRR267 runs showed 11.92–49.27% unique mapping and 47.75–85.41% unmapped-too-short reads, compared with 81.75–87.23% and 3.62–13.11%, respectively, for their matched SRR297 runs. To test whether the poor SRR267 mapping was confined to the start of the FASTQ files, 100,000 read pairs were sampled from each complete SRR267 dataset using seed 20260717. The random subsets reproduced the same pattern, with 12.23–50.33% unique mapping and 46.59–85.52% unmapped-too-short reads. The four SRR267 runs were therefore excluded in favour of their matched SRR297 runs and were not combined with them.

Raw paired-end FASTQ files were aligned without read trimming using STAR v2.7.11b (Dobin et al. 2013). The index used sjdbOverhang 158 and genomeSAsparseD 4. Alignment used two-pass Basic mode, local end alignment, BySJout filtering, a maximum of 20 mapping loci, one reported alignment per read, a maximum mismatch proportion of 0.04, minimum novel- and annotated-junction overhangs of 8 and 1 nucleotides, a minimum intron length of 20 nucleotides, maximum intron and mate-gap lengths of 1,000,000 nucleotides and MAPQ 60 for uniquely mapped reads. Additional settings were outSAMprimaryFlag OneBestScore, outSAMstrandField intronMotif, outSAMattributes NH HI AS nM MD and

quantMode GeneCounts. BAM files used for local sequence analysis were coordinate-sorted and indexed.

Gene-level abundance and Candidate C splice-chain analysis

Counts per million assigned fragments were calculated as the reverse-strand gene count divided by the total reverse-strand assigned fragments and multiplied by 10^6. Numerical values are reported in Table 1. STAR SJ.out.tab junctions were assigned to an annotated intron only when both genomic boundaries and strand orientation matched. Uniquely mapped and multimapping split-read counts were retained separately, and an intron absent from a run was assigned zero support.

Candidate C-specific transcription was assessed using the four introns of the five-exon transcript XM_038586400.1. Because Candidate C overlaps Candidate A on the opposite strand, interval-level coverage and Candidate C gene counts were not treated as transcript-specific evidence. Support required an exact Candidate C splice junction with the corresponding strand orientation.

Candidate A complete-chain junction audit

The six Candidate A models were grouped by identical coding sequences and translated products. Their annotated introns were collapsed by genomic start, genomic end and strand into 13 distinct junctions, designated J1–J13 in 5′-to-3′ transcript order. Coordinates and model membership are reported in Supplementary Table S4. A model-by-junction matrix was generated from strand-compatible STAR SJ.out.tab records across the eight pooled PBMC units.

For each intron contained in a public transcript, the matrix reports the summed uniquely mapped split-read count and the number of units with non-zero support. Junctions absent from a transcript were treated as structurally inapplicable rather than as zero-support

observations. Models encoding complete PYD–HIN proteins were ranked first by detection of every annotated intron in all eight units and then by the lowest summed unique-read support among their constituent introns. X5 and X6 satisfied the all-unit criterion; X5 was selected because its weakest supported junction had more uniquely mapped reads than the weakest X6 junction. The complete matrix is shown in Supplementary Fig. S1. Junction counts were not used to estimate isoform abundance or to phase the complete exon chain on one RNA molecule.

Candidate B transcript and coding-sequence reconstruction

Candidate B transcript models were compared across the five assemblies. The upstream first exon represented in ROS_Cfam_1.0 transcript XM_038448734.1 was aligned to the transcript-oriented AR106 genomic sequence, and its projected exon 1–exon 2 junction was queried in the eight STAR junction files. This upstream transcript structure was evaluated independently of the local coding-sequence discrepancy.

The assembly-specific proteins were XP_545739.2 from Dog10K_Boxer_Tasha, XP_038442327.1 from UU_Cfam_GSD_1.0, XP_038378526.1 from UMICH_Zoey_3.1, XP_038506745.1 from UNSW_CanFamBas_1.0 and XP_038304662.1 from ROS_Cfam_1.0. For the local coding analysis, the AR106 genomic exons annotated for XM_038586399.1 were concatenated in transcript orientation and compared with the distributed RefSeq transcript. BAM files were queried at NC_049259.1:23,783,787. Alignments were retained only when the NH tag was 1 and MAPQ was 60; alignments carrying unmapped, secondary, QC-failure, duplicate or supplementary flags were excluded using flag mask 0xF04. CIGAR operations were parsed at single-nucleotide resolution, and alignments were collapsed by SRA run and public read name so that each paired-end fragment was classified once.

Fragments were first classified according to retention of the focal genomic G, corresponding to C in transcript orientation, or support for the one-nucleotide gap in the public AR106

transcript. Downstream gap placement was evaluated only among fragments retaining the focal base and spanning NC_049259.1:23,783,787–23,783,795. Alternative transcript-oriented coding sequences were translated, checked for an initiation codon, internal stops and a terminal stop and compared with the five assembly-specific proteins. The PBMC-supported 441-amino-acid sequence encoding Ala37–Asn38 was used as the Candidate B representative. The upstream junction and local coding-site reconstruction were not considered phased across one complete native transcript.

Comparator selection and domain-resolved phylogenetic analysis

Comparator gene packages were retrieved through the NCBI Datasets v2 REST API. Queried GeneIDs were 9447, 3428, 4332 and 149628 for human; 100057148 and 106783181 for horse; 104971426 and 506759 for cattle; 100156073 and 146404038 for pig; 101095311 and 102902635 for cat; 112910073 for red fox; 100475349 for giant panda; 101670432 for ferret; 144307184 for golden jackal; 144228222 for spotted hyena; 114199140 for Steller sea lion; 122211631 for lion; 113932300 for California sea lion; 383619, 15951, 381308 and 236312 for mouse; and 111751912 and 100662402 for African elephant. Candidate A and Candidate B corresponded to canine GeneIDs 488622 and 111089951.

Exact duplicate protein sequences were collapsed before representative selection. Database symbols were used to locate candidate records but were not treated as evidence of orthology. Proteins were eligible for the expanded comparison only when one continuous predicted product contained complete specific PYD and HIN matches; eligible proteins were then prioritised by MANE Select status, curated NP accession and protein length. Incomplete MNDA-labelled products from spotted hyena and lion were excluded because they lacked complete specific PYD and HIN matches. Candidate C was excluded because it lacked a PYD and contained only an incomplete HIN-like match. Human IFI16, mouse Ifi204 and

California sea lion IFI16 each contributed two complete HIN domains, whereas the HIN-only cat protein XP_023102998.2 was retained only for HIN analysis.

CDD-bounded PYD and HIN sequences were aligned separately with MAFFT v7.526 using L-INS-i and the options --localpair --maxiterate 1000 --thread 4 --anysymbol (Katoh and Standley 2013). The expanded untrimmed alignments contained 25 PYD sequences over 74 sites and 29 HIN sequences over 174 sites. Primary maximum-likelihood inference used IQ-TREE v3.1.3 with -m MFP -B 1000 -alrt 1000 -bnni -T 4 -seed 20260727 (Wong et al. 2026). ModelFinder selected substitution models by the Bayesian information criterion (Kalyaanamoorthy et al. 2017; Minh et al. 2021), and retained JTT+G4 for PYD and Q.PLANT+F+G4 for HIN. Branch support was estimated using 1,000 SH-like approximate likelihood-ratio replicates and 1,000 ultrafast-bootstrap replicates with BNNI optimisation (Guindon et al. 2010; Hoang et al. 2018). Sensitivity analyses repeated inference on the untrimmed alignments using Q.MAMMAL+F+G4, JTT+G4, LG+G4 and WAG+G4 and repeated ModelFinder after removing alignment columns with less than 50% occupancy. The occupancy filter retained 74 PYD and 169 HIN sites. Recovery and support of the prespecified Candidate A and Candidate B focal groups were compared across these analyses; variation in deeper topology was not interpreted as evidence of stable orthology. Trees were inferred as unrooted estimates and midpoint rooted only for display.

Multiple-sequence-alignment visualisation

The 25-sequence PYD and 29-sequence HIN untrimmed alignments were ordered according to the displayed primary tree-tip order. Residues were recoded by physicochemical class only for Supplementary Figs. S2 and S3; phylogenetic inference used the original amino-acid states. Candidate A and Candidate B rows were outlined using the main-figure colours. The HIN alignment was divided into sites 1–87 and 88–174 solely for landscape-page layout.

For each column, the consensus row reports the modal non-gap residue, occupancy is the fraction of all sequences containing a residue and conservation is the frequency of the modal residue among non-gap sequences. These summaries were used for visualisation and did not filter or weight the phylogenetic analyses.



Fig. S1 Complete splice-junction audit of Candidate A public transcript models. The matrix shows the 13 distinct annotated introns (J1–J13, ordered from 5′ to 3′ in transcript orientation) represented among the six AR106 Candidate A protein-coding transcripts. For each intron contained in a model, cell fill encodes the number of the eight pooled PBMC analysis units with non-zero support; the first printed value is the summed uniquely mapped split-read count and the second is the number of supporting units expressed as *x*/8. Blank grey cells indicate that a junction is not part of the corresponding model and are structurally inapplicable rather than zero-support observations. The violet outline identifies X5, the selected 468-amino-acid representative. X5 and X6 had support for every annotated intron in all eight units, but the weakest X5 junction had 2,765 uniquely mapped split reads compared with 1,572 for X6; X5 also retained a complete PYD–HIN coding structure. X5 is therefore the most conservative representative among the evaluated public models, not evidence that it is the only native isoform. Junction evidence does not phase the complete exon chain on a single RNA molecule. Genomic coordinates and transcript membership for J1–J13 are provided in Supplementary Table S4.

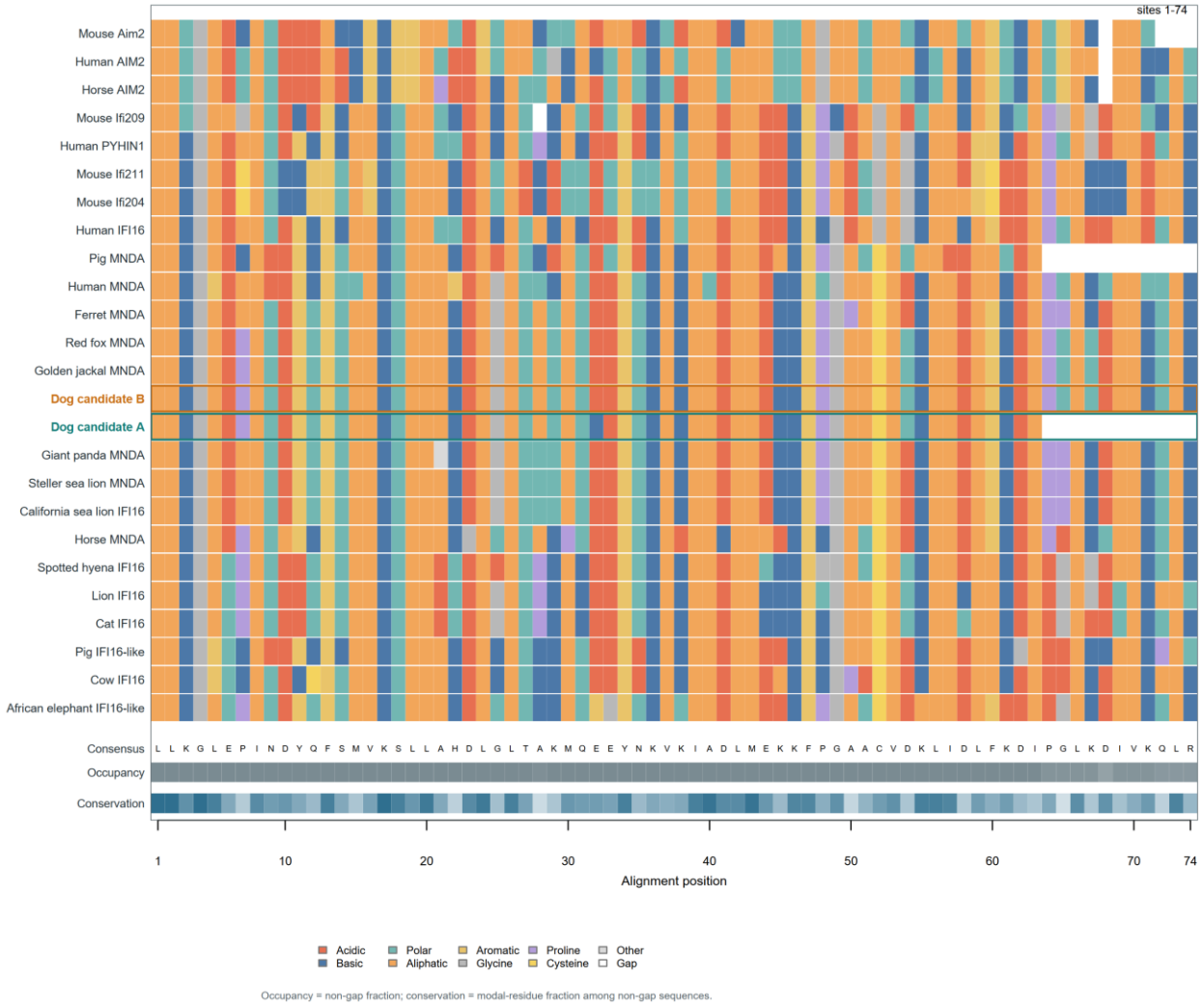


Fig. S2 PYD multiple-sequence-alignment overview used for phylogenetic analysis. Colour-coded overview of the expanded untrimmed CDD-bounded PYD alignment comprising 25 sequences and 74 aligned sites. Sequences are ordered as in the displayed primary PYD tree. Residues are coloured by physicochemical class, and gaps are white. Candidate A and Candidate B rows are outlined and labelled using the teal and orange mapping of the main figures. The consensus row gives the modal non-gap residue at each site. Occupancy is the non-gap fraction across all sequences, and conservation is the modal-residue fraction among non-gap sequences. Colour recoding and summary rows were used only for visualisation; maximum-likelihood analysis used the original amino-acid states. Sequence identifiers and CDD domain boundaries are listed in Supplementary Table S5.

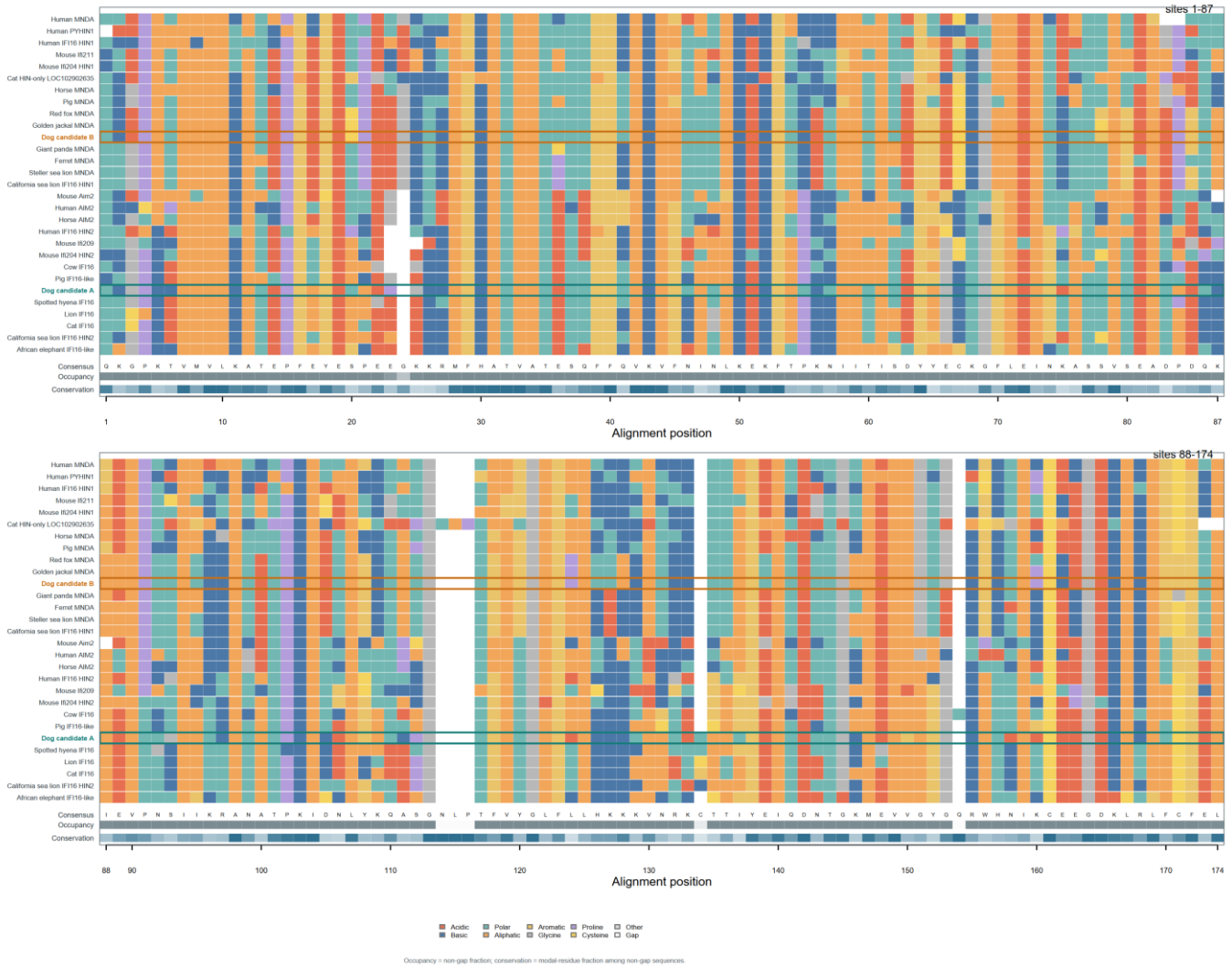


Fig. S3 HIN multiple-sequence-alignment overview used for phylogenetic analysis. Colour-coded overview of the expanded untrimmed CDD-bounded HIN alignment comprising 29 sequences and 174 aligned sites, divided into sites 1–87 and 88–174 for page layout. Sequences are ordered as in the displayed primary HIN tree. Residues are coloured by physicochemical class, and gaps are white. Candidate A and Candidate B rows are outlined and labelled using the teal and orange mapping of the main figures. The consensus row gives the modal non-gap residue at each site. Occupancy is the non-gap fraction across all sequences, and conservation is the modal-residue fraction among non-gap sequences. Colour recoding and summary rows were used only for visualisation; maximum-likelihood analysis used the original amino-acid states. Sequence identifiers and CDD domain boundaries are listed in Supplementary Table S5.

Table S1 Reference genomes and comparative locus records used in Fig. 1

Species	Reference assembly	Chromosome / annotation	Order	Displayed locus	GeneID	Representative protein or figure role
Human	GRCh38.p14 GCF_000001405.40	NC_000001.11 RS_2025_08	1	CADM3	57863	5′ flanking locus
			2	AIM2	9447	NP_004824.1
			3	IFI16	3428	NP_001363516.1
			4	PYDC5/POP3	107181291	NP_001306939.1
			5	PYHIN1	149628	NP_689714.2
			6	MNDA	4332	NP_002423.1
			7	SPTA1	6708	3′ flanking locus
Mouse	GRCm39 GCF_000001635.27	NC_000067.7 RS_2024_02	1	Cadm3	94332	5′ flanking locus
			2	Aim2	383619	Schematic PYHIN unit
			3	Ifi206	102639543	Schematic PYHIN unit
			4	Ifi214	545384	Schematic PYHIN unit
			5	Ifi213	623121	Schematic PYHIN unit
			6	Ifi209	236312	Schematic PYHIN unit
			7	Ifi208	100033459	Schematic PYHIN unit
			8	Ifi207	226691	Schematic PYHIN unit
			9	Ifi204	15951	Schematic PYHIN unit
			10	Mndal	100040462	Schematic PYHIN unit
			11	Ifi211	381308	Schematic PYHIN unit
			12	Ifi203	15950	Schematic PYHIN unit
			13	Ifi202b	26388	Schematic PYHIN unit
			14	Ifi205	226695	Schematic PYHIN unit
			15	Spta1	20739	3′ flanking locus
Dog	UU_Cfam_GSD_1.0 GCF_011100685.1	NC_049259.1 AR106	1	CADM3	610435	5′ flanking locus
			2	Candidate C / LOC111094466	111094466	XP_038442328.1
			3	Candidate A / LOC488622	488622	XP_038442325.1
			4	Candidate B / current MNDA	111089951	RNA-supported AN441
			5	SPTA1	488616	3′ flanking locus

Rows are ordered as displayed in the comparative schematic. Human and mouse genomic distances were schematic; dog loci were positioned proportionally in Fig. 1. Mouse protein accessions were not used to scale the schematic.

Table S2 Cross-assembly audit of the canine PYHIN candidate records

Assembly	Chromosome record	Candidate A	Candidate B	Candidate C	Candidate C domain audit	Intact AIM2 model
UU_Cfam_GSD_1.0 GCF_011100685.1	NC_049259.1	6 models 412/468/509 aa complete PYD–HIN	1 model 441 aa complete PYD–HIN	1 model 187 aa	no PYD partial HIN in 1/1 model	not supported
ROS_Cfam_1.0 GCF_014441545.1	NC_051842.1	6 models 412/468/509 aa complete PYD–HIN	1 model 441 aa complete PYD–HIN	1 model 187 aa	no PYD partial HIN in 1/1 model	not supported
UMICH_Zoey_3.1 GCF_005444595.1	NC_049298.1	5 models 412/468/509 aa complete PYD–HIN	1 model 441 aa complete PYD–HIN	1 model 187 aa	no PYD partial HIN in 1/1 model	not supported
UNSW_CanFamBas_1.0 GCF_013276365.1	NC_049779.1	5 models 412/468/509 aa complete PYD–HIN	1 model 441 aa complete PYD–HIN	1 model 187 aa	no PYD partial HIN in 1/1 model	not supported
Dog10K_Boxer_Tasha GCF_000002285.5	NC_006620.4	6 models 412/468/509 aa complete PYD–HIN	1 model 441 aa complete PYD–HIN	3 models 187/190/191 aa	no PYD partial HIN in 1/3 models	not supported

A complete PYHIN model required one continuous predicted protein containing complete specific PYD and HIN domain matches. Candidate C did not satisfy this criterion in any evaluated assembly.

Table S3 Public canine PBMC RNA-sequencing units and run-selection audit

Breed	Condition	Retained run	BioSample	Retained experiment	Excluded earlier run	Excluded experiment	Run-selection note
Gaddi	Unstimulated	SRR28408367	SAMN40563700	SRX24012849	—	—	Retained pooled unit
Gaddi	LPS	SRR28408365	SAMN40563702	SRX24012851	—	—	Retained pooled unit
Gaddi	Poly(I:C)	SRR28408366	SAMN40563701	SRX24012850	—	—	Retained pooled unit
Gaddi	CpG ODN	SRR28408364	SAMN40563703	SRX24012852	—	—	Retained pooled unit
Labrador	Unstimulated	SRR29708343	SAMN38085041	SRX25210785	SRR26719004	SRX22417956	SRR297 retained after paired-end mapping review
Labrador	LPS	SRR29708345	SAMN38085043	SRX25210783	SRR26719002	SRX22417958	SRR297 retained after paired-end mapping review
Labrador	Poly(I:C)	SRR29708344	SAMN38085042	SRX25210784	SRR26719003	SRX22417957	SRR297 retained after paired-end mapping review
Labrador	CpG ODN	SRR29708346	SAMN38085044	SRX25210782	SRR26719001	SRX22417959	SRR297 retained after paired-end mapping review

Each breed-condition combination represents one pooled descriptive unit. Earlier Labrador runs were excluded after paired-end mapping review and were not combined with the retained SRR297 runs.

Table S4 Candidate A junction-coordinate key

Junction ID	AR106 genomic interval	Strand	Candidate A models
J1	NC_049259.1:23746887–23768643	-	X4
J2	NC_049259.1:23746887–23768553	-	X3
J3	NC_049259.1:23746887–23747206	-	X2
J4	NC_049259.1:23742244–23746767	-	X1, X2, X3, X4, X5, X6
J5	NC_049259.1:23741051–23741960	-	X1, X2, X3, X4, X5, X6
J6	NC_049259.1:23739930–23740919	-	X1, X2, X3, X4, X5, X6
J7	NC_049259.1:23739232–23739761	-	X1, X2, X3, X4
J8	NC_049259.1:23738274–23739761	-	X5
J9	NC_049259.1:23738274–23739108	-	X1, X2, X3, X4
J10	NC_049259.1:23732182–23739761	-	X6
J11	NC_049259.1:23732182–23738105	-	X1, X2, X3, X4, X5
J12	NC_049259.1:23730384–23731761	-	X1, X2, X3, X4, X5, X6
J13	NC_049259.1:23729235–23730197	-	X1, X2, X3, X4, X5, X6

Junctions are listed in transcript orientation. Model membership refers to the six AR106 Candidate A transcript models evaluated in Supplementary Fig. S1.

Table S5 Sequence metadata and CDD-bounded intervals used for phylogenetic analysis

Tree label	Species	Gene symbol	Sequence identifier	Domain	Residues (aa)
GiantPanda_MNDA_PYD	*Ailuropoda melanoleuca*	MNDA	XP_034523221.1	PYD	10–83
Cow_IFI16_PYD	*Bos taurus*	IFI16	XP_002685923.1	PYD	10–83
GoldenJackal_MNDA_PYD	*Canis aureus*	MNDA	XP_077742852.1	PYD	10–83
Dog_Candidate_A_PYD	*Canis lupus familiaris*	Candidate_A_LOC488622	XP_038442325.1	PYD	10–72
Dog_Candidate_B_PYD	*Canis lupus familiaris*	Candidate_B_MNDA	Candidate_B_AN441	PYD	10–83
SpottedHyena_IFI16_PYD	*Crocuta crocuta*	IFI16	XP_077616153.1	PYD	10–83
Horse_AIM2_PYD	*Equus caballus*	AIM2	XP_014595286.1	PYD	10–82
Horse_MNDA_PYD	*Equus caballus*	MNDA	XP_014595284.3	PYD	10–83
StellerSeaLion_MNDA_PYD	*Eumetopias jubatus*	MNDA	XP_027946735.1	PYD	10–83
Cat_IFI16_PYD	*Felis catus*	IFI16	XP_044904507.1	PYD	10–83
Human_AIM2_PYD	*Homo sapiens*	AIM2	NP_004824.1	PYD	10–82
Human_IFI16_PYD	*Homo sapiens*	IFI16	NP_001363516.1	PYD	10–83
Human_MNDA_PYD	*Homo sapiens*	MNDA	NP_002423.1	PYD	10–83
Human_PYHIN1_PYD	*Homo sapiens*	PYHIN1	NP_689714.2	PYD	10–83
AfricanElephant_IFI16like_PYD	*Loxodonta africana*	LOC100662402	XP_023409917.1	PYD	10–83
Mouse_Aim2_PYD	*Mus musculus*	Aim2	NP_001013801.2	PYD	10–79
Mouse_Ifi204_PYD	*Mus musculus*	Ifi204	NP_032355.2	PYD	10–83
Mouse_Ifi209_PYD	*Mus musculus*	Ifi209	NP_778191.1	PYD	10–82
Mouse_Ifi211_PYD	*Mus musculus*	Ifi211	NP_001028622.1	PYD	10–83
Ferret_MNDA_PYD	*Mustela putorius furo*	MNDA	XP_004775868.1	PYD	10–83
Lion_IFI16_PYD	*Panthera leo*	IFI16	XP_042780847.1	PYD	10–83
Pig_IFI16like_PYD	*Sus scrofa*	LOC100156073	XP_084877370.1	PYD	10–83
Pig_MNDA_PYD	*Sus scrofa*	MNDA	XP_084877381.1	PYD	10–72
RedFox_MNDA_PYD	*Vulpes vulpes*	MNDA	XP_072615870.1	PYD	10–83
CaliforniaSeaLion_IFI16_PYD	*Zalophus californianus*	IFI16	XP_035578069.1	PYD	10–83
GiantPanda_MNDA_HIN	*Ailuropoda melanoleuca*	MNDA	XP_034523221.1	HIN	207–375
Cow_IFI16_HIN	*Bos taurus*	IFI16	XP_002685923.1	HIN	324–491
GoldenJackal_MNDA_HIN	*Canis aureus*	MNDA	XP_077742852.1	HIN	211–379
Dog_Candidate_A_HIN	*Canis lupus familiaris*	Candidate_A_LOC488622	XP_038442325.1	HIN	262–430
Dog_Candidate_B_HIN	*Canis lupus familiaris*	Candidate_B_MNDA	Candidate_B_AN441	HIN	211–379
SpottedHyena_IFI16_HIN	*Crocuta crocuta*	IFI16	XP_077616153.1	HIN	470–638
Horse_AIM2_HIN	*Equus caballus*	AIM2	XP_014595286.1	HIN	154–321
Horse_MNDA_HIN	*Equus caballus*	MNDA	XP_014595284.3	HIN	208–376
StellerSeaLion_MNDA_HIN	*Eumetopias jubatus*	MNDA	XP_027946735.1	HIN	206–374
Cat_HINonly_LOC102902635_HIN	*Felis catus*	LOC102902635	XP_023102998.2	HIN	15–184
Cat_IFI16_HIN	*Felis catus*	IFI16	XP_044904507.1	HIN	486–654
Human_AIM2_HIN	*Homo sapiens*	AIM2	NP_004824.1	HIN	150–317
Human_IFI16_HIN1	*Homo sapiens*	IFI16	NP_001363516.1	HIN1	201–369
Human_IFI16_HIN2	*Homo sapiens*	IFI16	NP_001363516.1	HIN2	574–740
Human_MNDA_HIN	*Homo sapiens*	MNDA	NP_002423.1	HIN	208–374
Human_PYHIN1_HIN	*Homo sapiens*	PYHIN1	NP_689714.2	HIN	212–379
AfricanElephant_IFI16like_HIN	*Loxodonta africana*	LOC100662402	XP_023409917.1	HIN	374–541
Mouse_Aim2_HIN	*Mus musculus*	Aim2	NP_001013801.2	HIN	156–321
Mouse_Ifi204_HIN1	*Mus musculus*	Ifi204	NP_032355.2	HIN1	225–393
Mouse_Ifi204_HIN2	*Mus musculus*	Ifi204	NP_032355.2	HIN2	429–595
Mouse_Ifi209_HIN	*Mus musculus*	Ifi209	NP_778191.1	HIN	231–396
Mouse_Ifi211_HIN	*Mus musculus*	Ifi211	NP_001028622.1	HIN	225–393
Ferret_MNDA_HIN	*Mustela putorius furo*	MNDA	XP_004775868.1	HIN	212–380
Lion_IFI16_HIN	*Panthera leo*	IFI16	XP_042780847.1	HIN	436–604
Pig_IFI16like_HIN	*Sus scrofa*	LOC100156073	XP_084877370.1	HIN	357–524
Pig_MNDA_HIN	*Sus scrofa*	MNDA	XP_084877381.1	HIN	208–376
RedFox_MNDA_HIN	*Vulpes vulpes*	MNDA	XP_072615870.1	HIN	224–392
CaliforniaSeaLion_IFI16_HIN1	*Zalophus californianus*	IFI16	XP_035578069.1	HIN1	206–374
CaliforniaSeaLion_IFI16_HIN2	*Zalophus californianus*	IFI16	XP_035578069.1	HIN2	578–746

PYD and HIN domains were aligned and analysed separately. Human IFI16, mouse Ifi204 and California sea lion IFI16 each contributed two HIN domains where complete CDD-bounded matches were available.