

Molecular Systematics & Complexity | Signal Rating: 1/5 Stars | Oct 28, 2026 | 12 min read

Carnivora Mitochondrial COI DNA Barcoding Cluster Analysis & Information-Theoretic Codon Complexity Across 16 Reference Mitogenomes

Multi-threaded Kimura 2-Parameter (K2P) and Tamura-Nei (TN93) phylogenetics, Automatic Barcode Gap Discovery (ABGD), and Lempel-Ziv / Shannon codon entropy profiling across Ursidae, Felidae, Canidae, and Mustelidae.

1. Abstract & Worker-Agent Pipeline Architecture

Mitochondrial Cytochrome c Oxidase Subunit I (COX1 / COI) serves as the universal metazoan DNA barcode locus because its maternal inheritance, absence of intronic recombination, and rapid synonymous substitution rate cleanly separate intraspecific polymorphism from interspecific divergence [1] . However, classical barcoding pipelines stop at pairwise Kimura 2-Parameter (K2P) distance tables [2] , ignoring how thermodynamic protein-folding constraints in complex IV of the oxidative phosphorylation chain shape the underlying nucleotide symbol string. Using the standalone Rust bio-suite v0.1.0 engine in CRV Science Labs , we chained four automated stages into a single deterministic worker-agent execution (coi-barcode-cluster-v1 , completing in 10-15 ms from local SHA-256 cache across all CPU threads via rayon):

* Tier 1 Polite Acquisition & Frame Verification (bio-suite fetch): Retrieved 16 complete, in-frame 1,545 bp (515 codon) vertebrate mitochondrial COX1 sequences from NCBI RefSeq/GenBank spanning 4 Carnivora families -- Ursidae (*Ursus arctos* , *Ursus maritimus* , *Ursus americanus* , *Ursus thibetanus*), Felidae (*Panthera tigris* , *Panthera pardus* , *Felis catus*), Canidae (*Canis lupus* , *Canis lupus familiaris* , *Vulpes vulpes*), and Mustelidae (*Meles anakuma* , *Mustela putorius*) [6] .

* Tier 2 Codon-Anchored MSA & Distance Matrix (bio-suite phylo): Aligned all 16 coding frames across 1,545 bp without frameshifts, identifying 970 strictly conserved sites (62.78%), 575 variable sites (37.22%), and 514 parsimony-informative sites (33.27%), followed by all 120 pairwise p-distance, Kimura 2-Parameter (K2P) [2] , and Tamura-Nei (TN93) [3] substitution calculations.

* Tier 2 Neighbor-Joining & ABGD Cluster Delineation (bio-suite phylo): Reconstructed the unrooted Neighbor-Joining (NJ) phylogeny [4] with 500 parallel bootstrap replicates and partitioned specimens into Molecular Operational Taxonomic Units (MOTUs) at the 2.5% (K2P = 0.025) vertebrate barcoding threshold [5] .

* Tier 4 Information-Theoretic & Fractal Overlay (bio-suite complexity): Computed k -mer Shannon block entropies (H_1 , H_2 , H_3), codon-position entropies (H_{pos1} , H_{pos2} , H_{pos3}), GC3% , CpG observed/expected ratios, Lempel-Ziv 1976 (LZ76) algorithmic compressibility, and Purine- Pyrimidine DNA walk Detrended Fluctuation Analysis (alpha DFA) across every clade.

2. K2P Divergence & Clade Bootstrap Summary

Across the 120 pairwise comparisons, evolutionary divergence scales monotonically across the Linnaean hierarchy: mean intraspecific K2P = 0.0091 (0.91%), mean intrageneric K2P = 0.0813 (8.13%), mean intrafamily K2P = 0.1641 (16.41%), and mean interfamilial K2P = 0.2316 (23.16%), with a panel-wide transition/transversion ratio of Ti/Tv = 6.9731 . Click Copy for Excel (TSV) to paste this summary table directly into Excel or R:

Family	Clade	Common Group	Sampled Taxa (n)	Species (n)	Mean Within-Clade K2P	Bootstrap Support (%)	Diagnostic Synapomorphies	Mean GC (%)	Mean GC3 (%)	Mean CpG (O/E)	Mean H pos3 (bits)	Mean Norm LZ76
Ursidae	Bears	7	4	0.0691	98.8	28	42.52	39.24	0.5331	1.9371	0.9737	
Felidae	Cats	4	3	0.1086	91.4	27	42.33	36.50	0.5810	1.9079	0.9753	
Canidae	Canids	3	2	0.1165	65.4	28	41.32	33.98	0.6388	1.8938	0.9725	
Mustelidae	Mustelids	2	2	0.2054	21.2	21	40.71	33.11	0.5567	1.8841	0.9719	

3. ABGD MOTU Clustering & Neighbor-Joining Phylogeny

Single-linkage Automatic Barcode Gap Discovery (ABGD) at the K2P = 0.025 threshold partitions the 16 specimens into 11 MOTUs , highlighting three baseline clustering behaviors:

* MOTU-01 -- Pleistocene Peripatric Speciation & Mitogenome Capture in *Ursus arctos* / *Ursus maritimus* : Eurasian and Hokkaido brown bears (*U. arctos* , intraspecific K2P = 0.0124) and polar bears (*U. maritimus* , intraspecific K2P = 0.0000 between RefSeq NC_003428.1 and isolate GU573490.1) exhibit an interspecific divergence of only K2P = 0.0171-0.0191 (23 transitions, 3 transversions, Ti/Tv = 7.67), coalescing into MOTU-01 (nearest outgroup neighbor: *Ursus thibetanus* at K2P = 0.0950). This reproduces the well-documented Late Pleistocene mitochondrial paraphyly between brown and polar bears [7] .

* MOTU-02 & MOTU-03 -- Glacial Refugial Divergence in *Ursus americanus* : The two American black bear mitogenomes (NC_003426.1 vs. isolate JX196366.1) diverge by K2P = 0.0300 (3.00%), exceeding the 2.5% threshold and splitting into MOTU-02 and MOTU-03 across Eastern and Western forest refugia.

* MOTU-08 -- Holocene Domestication Continuity in *Canis lupus* : Gray wolf (NC_008092.1) and domestic dog (NC_002008.4) differ by 5 transitions and 0 transversions across 1,545 bp (K2P = 0.0032), uniting into MOTU-08 separated from red fox (*Vulpes vulpes* , MOTU-09, K2P = 0.1718).

Below are both the visual ASCII Neighbor-Joining Cladogram (showing K2P branch lengths, bootstrap support, and ABGD MOTU assignments) and the raw Newick Topology Codebox reconstructed by bio-suite phylo :

4. Sequence Dataset & Complexity Scores

Overlaying Tier-4 complexity metrics onto the 16 COI coding sequences reveals how mitochondrial purifying selection and mutational bias interact across the 515 codons . Click Copy for Excel (TSV) to copy all 16 accession numbers and atomic numeric columns directly into a spreadsheet:

| Accession Slice || Species || Common Name || Family || GC (%) || GC3 (%) || CpG (O/E) || H pos1 (bits) || H pos2 (bits) || H pos3 (bits) || Shannon H3 (bits) || LZ76 Phrases || Norm LZ76 || DFA Alpha || Hurst H |

| NC_003427.1:6484-8028 || Ursus arctos || Eurasian brown bear || Ursidae || 43.04 || 38.45 || 0.5520 || 1.9870 || 1.8922 || 1.9322 || 5.8526 || 283 || 0.9702 || 0.4882 || 0.4926 |

| AP012578.1:6392-7936 || Ursus arctos || Hokkaido brown bear || Ursidae || 43.17 || 39.22 || 0.5631 || 1.9877 || 1.8884 || 1.9375 || 5.8632 || 284 || 0.9736 || 0.4907 || 0.4983 |

| NC_003428.1:6479-8023 || Ursus maritimus || Polar bear RefSeq || Ursidae || 42.78 || 38.06 || 0.5581 || 1.9870 || 1.8903 || 1.9333 || 5.8522 || 290 || 0.9942 || 0.4882 || 0.4926 |

| GU573490.1:6272-7816 || Ursus maritimus || Polar bear 495 || Ursidae || 42.78 || 38.06 || 0.5581 || 1.9870 || 1.8903 || 1.9333 || 5.8522 || 290 || 0.9942 || 0.4882 || 0.4926 |

| NC_003426.1:6291-7835 || Ursus americanus || Black bear RefSeq || Ursidae || 41.62 || 34.95 || 0.4830 || 1.9868 || 1.8903 || 1.9130 || 5.8326 || 280 || 0.9599 || 0.4937 || 0.4933 |

| JX196366.1:5883-7427 || Ursus americanus || Isolate BLK014 || Ursidae || 42.07 || 35.92 || 0.5162 || 1.9870 || 1.8903 || 1.9196 || 5.8384 || 279 || 0.9565 || 0.4860 || 0.4878 |

| NC_009971.1:6243-7787 || Ursus thibetanus || Asian black bear || Ursidae || 42.14 || 36.12 || 0.5008 || 1.9877 || 1.8877 || 1.9228 || 5.8441 || 282 || 0.9668 || 0.4821 || 0.4899 |

| NC_010642.1:6280-7824 || Panthera tigris || Tiger RefSeq || Felidae || 42.98 || 37.86 || 0.6245 || 1.9875 || 1.8916 || 1.9274 || 5.8696 || 284 || 0.9736 || 0.5122 || 0.5090 |

| EF551003.1:6280-7824 || Panthera tigris || Amur tiger || Felidae || 42.98 || 37.86 || 0.6245 || 1.9875 || 1.8916 || 1.9274 || 5.8696 || 284 || 0.9736 || 0.5122 || 0.5090 |

| NC_010641.1:6305-7849 || Panthera pardus || Leopard RefSeq || Felidae || 41.68 || 34.56 || 0.5436 || 1.9861 || 1.8909 || 1.8915 || 5.8419 || 288 || 0.9873 || 0.5121 || 0.5137 |

| NC_001700.1:6216-7760 || Felis catus || Domestic cat RefSeq || Felidae || 41.68 || 35.73 || 0.5315 || 1.9849 || 1.8903 || 1.8854 || 5.8521 || 282 || 0.9668 || 0.5191 || 0.5128 |

NC_008092.1:5350-6894		Canis lupus		Gray wolf RefSeq		Canidae		40.71						
32.82		0.6666		1.9846		1.8937		1.8716		5.8386		284		0.9736
0.5029		0.5120												

| NC_002008.4:5349-6893 || Canis lupus familiaris || Domestic dog || Canidae ||
40.78 || 32.82 || 0.6800 || 1.9849 || 1.8937 || 1.8726 || 5.8422 || 284 ||
0.9736 || 0.5015 || 0.5112 |

NC_008434.1:5350-6894		Vulpes vulpes		Red fox RefSeq		Canidae		42.46						
36.31		0.5698		1.9859		1.8920		1.9371		5.8690		283		0.9702
0.5046		0.5161												

| NC_009677.1:5339-6883 || Meles anakuma || Japanese badger || Mustelidae ||
40.26 || 31.46 || 0.5162 || 1.9838 || 1.8920 || 1.8548 || 5.8229 || 282 ||
0.9668 || 0.4910 || 0.5050 |

| NC_020638.1:5343-6887 || Mustela putorius || European polecat || Mustelidae ||
41.17 || 34.76 || 0.5973 || 1.9855 || 1.8909 || 1.9134 || 5.8501 || 285 ||
0.9770 || 0.5105 || 0.5128 |

5. Codon Entropy Patterns & 1* Signal Limitations

Overlaying the complexity metrics onto the 16 COI sequences illustrates standard features of vertebrate mitochondrial coding sequences:

* Second-Position Entropy Constraint ($H_{pos2} \sim 1.890$ bits): In the vertebrate mitochondrial genetic code, the second nucleotide of a codon determines amino-acid hydrophobicity (uracil/ thymine at position 2 encodes hydrophobic residues -- Phe, Leu, Ile, Met, Val -- across the 12 transmembrane alpha-helices of COX1). Purifying selection keeps H_{pos2} within 1.8877-1.8937 bits across the sampled Carnivora taxa during evolution , whereas 3rd-position wobble bases vary in GC 3 content (31.46% in Meles anakuma to 39.22% in Hokkaido Ursus arctos) via synonymous transition mutations ($Ti/Tv = 6.97$).

* CpG Depletion & Coding Compressibility: Although mitochondrial CpG dinucleotides are depleted across all 16 mitogenomes ($CpG\ O/E = 0.4830$ - 0.6800), 3-mer Shannon entropy ($H_3 = 5.8229$ - 5.8696 bits) and normalized Lempel-Ziv complexity ($LZ76 = 0.9565$ - 0.9942) remain uniform, and DNA walk Detrended Fluctuation Analysis stays near an uncorrelated Brownian exponent ($\alpha_{DFA} = 0.4821$ - 0.5191) expected for intronless coding sequences.

Signal Rating Assessment (1* Baseline / Review Note): This report is rated 1* because it serves as a small-sample ($n = 16$ sequences, single 1,545 bp mitochondrial locus) pipeline benchmark for bio-suite . While it verifies that K2P/NJ clustering and codon-position entropy calculations run accurately and flags expected taxonomic boundary cases (such as Ursus arctos / U. maritimus mitochondrial paraphyly), the dataset is too small and locus-restricted to link sequence complexity variations to any novel biological or physiological factor.

References & Primary Sources

* [1] Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B:

Biological Sciences, 270(1512), 313-321. <https://doi.org/10.1098/rspb.2002.2218>

* [2] Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2), 111-120. <https://doi.org/10.1007/BF01731581>

* [3] Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10(3), 512-526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>

* [4] Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>

* [5] Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864-1877. <https://doi.org/10.1111/j.1365-294X.2011.05239.x>

* [6] NCBI RefSeq Carnivora Mitochondrial Genomes Panel (NC_003427.1, AP012578.1, NC_003428.1, GU573490.1, NC_003426.1, JX196366.1, NC_009971.1, NC_010642.1, EF551003.1, NC_010641.1, NC_001700.1, NC_008092.1, NC_002008.4, NC_008434.1, NC_009677.1, NC_020638.1). https://www.ncbi.nlm.nih.gov/nuccore/NC_003427.1

* [7] Hailer, F., Kutschera, V. E., Hallstr m, B. M., Klassert, D., Fain, S. R., Leonard, J. A., Arnason, U., & Janke, A. (2012). Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. *Science*, 336(6079), 344-347. <https://doi.org/10.1126/science.1216424>