Eurasia Specialized Veterinary Publication

International Journal of Veterinary Research and Allied Science
ISSN:3062-357X

2022, Volume 2, Issue 2, Page No: 78-88

Available online at: www.esvpub.com/

Domestication and Divergent Evolution of Neuropeptide Prohormone Genes in Cetartiodactyla: Insights from 118 Genomes

Mikkel Jensen[1], Emma Nielsen[1*], Lucas Hansen[1]

[1]Section for Cell and Neurobiology, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100, Copenhagen, Denmark.

***E-mail** ✉ emma.ni1sen@outlook.com

ABSTRACT

The evolutionary and domestication processes have shaped the sequences of neuropeptide prohormone genes that mediate cell–cell communication, affecting a wide range of biological systems dependent on neuropeptide signaling. Understanding these modifications helps explain physiological contrasts between domesticated Cetartiodactyla, including cattle, pigs, and llamas, and their wild counterparts such as hippopotamuses, giraffes, and whales. A systematic investigation was carried out to assess how evolutionary and domestication pressures influenced neuropeptide prohormone proteins, which are precursors of active neuropeptides. Genomic data from 118 Cetartiodactyla species, representing 22 families, were screened for **98** neuropeptide prohormone genes. Among suborders, Ruminantia retained PYY2 while losing RLN1. Alterations in the sequences of GNRH2, IAPP, INSL6, POMC, PRLH**,** and TAC4 may have led to the absence of certain functional neuropeptides in some lineages. Evolutionary modeling indicated that most neuropeptide prohormone genes tend to resist amino acid substitutions introducing bulky or hydrophobic residues. Notably, variations distinguishing wild and domestic species were linked to molecular mechanisms involved in "fight or flight" regulation. Overall, the findings emphasize the necessity of comparing neuropeptide prohormone gene repertoires among both closely and distantly related species. These results strengthen the groundwork for further functional studies of the neuropeptidome relevant to health, behavioral adaptation, and animal production.

Keywords: Neuropeptide, Prohormone, Domestication, Cetartiodactyla, Evolution

Received: 01 August 2022
Revised: 19 October 2022
Accepted: 21 October 2022

How to Cite This Article: Jensen M, Nielsen E, Hansen L. Domestication and Divergent Evolution of Neuropeptide Prohormone Genes in Cetartiodactyla: Insights from 118 Genomes. Int J Vet Res Allied Sci. 2022;2(2):78-88. https://doi.org/10.51847/WX16oPLKkE

Introduction

Domestication generally involves isolating species from their natural ecosystems and introducing behavioral and physiological adaptations through artificial selection and controlled environments [1]. This phenomenon is exemplified by the mammalian superorder Cetartiodactyla**,** which includes both domesticated livestock and wild species thriving across alpine, desert, and aquatic habitats, managed under varied human systems including precision farming [2-4]. Of the four Cetartiodactyla suborders, three contain domesticated species: Ruminantia (cattle, sheep, goats, deer), Suina (pigs, peccaries), and Tylopoda (camels, llamas, alpacas). These terrestrial domesticates serve multiple purposes such as food, fiber, transport, companionship, and display. The fourth suborder, Whippomorpha (hippopotamuses and whales), remains primarily wild, though individuals are maintained in zoological and research settings for conservation [5].

Livestock domestication began around 11,000 years ago, starting with cattle (*Bos taurus*), goats (*Capra hircus*), sheep (*Ovis aries*), and pigs (*Sus scrofa*) [2-4]. The Tylopoda group underwent domestication between 3000 and

7000 years ago [6], while reindeer (*Rangifer tarandus*) remain semi-domesticated with both free-ranging and managed herds [7]. The domestication process alters biological pathways governing aggression, fear, and escape behaviors, and additionally influences functions related to fertility, metabolism, development, growth, feeding, reproduction, and disease resistance via selection and breeding [1-4].

Artificial selection has driven population-level genetic variation, thereby modifying physiology and behavior in domestic animals. These biological changes are closely tied to intercellular signaling mechanisms involving neuropeptides that act as chemical messengers by binding to prohormone receptors. Hormones such as adrenocorticotropic hormone (from POMC), oxytocin (from OXT), arginine vasopressin (AVP), and corticotropin-releasing hormone (CRH) regulate aggression, social bonding, and stress [8]. Meanwhile, appetite and feeding behaviors are governed by insulin-like growth factor 2 (IGF2), neuropeptide Y (NPY), and ghrelin (derived from GHRL).

Evolutionary mechanisms—including mutation, selection, gene flow, and drift—have also contributed to the diversification of neuropeptide sequences and their associated physiological traits. Cross-species comparisons of GCG and GIP genes provide insight into these evolutionary effects [9]. A mutation in the MC1R receptor, which binds POMC peptides, plays a significant role in coat color variation among wild and domestic species [10]. Though natural and artificial selection often overlap, the selective pressures acting on domestic versus wild populations differ markedly [11]. Analysis of the melanocortin pathway revealed that MC1R exerted the strongest coevolutionary effect, while POMC and AGRP exhibited the weakest and third-weakest influences, respectively [12].

Predicting domestication impacts on neuropeptides is challenging since these molecules arise from intricate post-translational processing rather than direct gene transcription. Neuropeptide prohormone genes encode precursor proteins that include a signal peptide directing secretion. Following signal peptide removal [13], the protein undergoes cleavage at furin-like recognition sites [14], typically characterized by arginine or lysine residues at or near the cut site. Further post-translational modifications yield bioactive neuropeptides.

A specialized bioinformatics framework capable of analyzing protein sequences and modeling this complex processing pathway has been developed [15, 16]. Using this method, researchers have identified neuropeptide prohormone genes that were either lost or duplicated through evolutionary events [17-19]. The current study aimed to explore sequence variation in neuropeptide prohormone genes resulting from domestication and evolutionary forces. To this end, a comprehensive genomic dataset covering both wild and domestic Cetartiodactyla species was compiled. This dataset enabled the identification of gene gains, losses, and hybridization events occurring during evolutionary and domestication histories [20-22].

Materials and Methods

Genomic assemblies from 118 members of the order Cetartiodactyla, covering four suborders and 22 distinct families, were obtained from the National Center for Biotechnology Information (NCBI) database [23]. Details of the analyzed taxa are summarized in **Table 1**. Within this dataset, Ruminantia accounted for 82 species—including cattle, sheep, goats, and deer—distributed among six families. The Suina group was composed of two species (pigs and peccaries), while Tylopoda contained seven species belonging to two tribes (camels and llamas). The Whippomorpha lineage included 27 species, such as whales and the hippopotamus.

Table 1. Overview of wild and domesticated Cetartiodactyla species by taxonomic classification.

Suborder and IF or Fam[1]	N[2]	Parvorder/Subfamily/Tribe/Genus[3]	D	W
Ruminantia				
Bovidae	7	Bovinae: 4,12; Caprinae: 2,12; Alcelaphinae: 0,4; Reduncinae: 0,3; Cephalophinae: 0,3; Hippotraginae: 0,4; Antilopinae: 0,12; Aepycerotinae: 0,1		
Cervidae	1	Muntiacinae: 0,3; Cervinae: 0,5; Hydropotinae: 0,1; Odocoileinae: 1,6		
Other	0	Moschidae: 0,3; Tragulidae: 0,2; Giraffidae: 0,3; Antilocapridae: 0,1		
Suina	1	Sus: 1,0; Catagonus: 0,1		
Tylopoda				

Camelidae	4	Camelini: 2,1; Lamini: 2,2	
Whippomorpha	0	Hippopotamidae: 0,1; Odontoceti: 0,19; Mysticeti: 0,7	
Total	13		105

[1]Cetartiodactyla suborder and infraorder (IF) or family (Fam);
[2]number of domesticated (D) and wild (W) species within each infraorder or family;
[3]number of domesticated species followed by wild species in each parvorder, subfamily, or tribe.

The analyzed species were categorized according to suborder, family, subfamily, and domestication status. Altogether, 12 taxa were recognized as *domesticated* and 106 taxa as *wild*. The wild Bactrian camel (*Camelus ferus*) was excluded from either group because its domestication history for the sequenced individuals remains unresolved [24]. Wild representatives from Ruminantia, Suina, and Tylopoda were classified as *terrestrial non-Cetacea* (a total of 91 species), whereas aquatic taxa consisted of the Cetacea members of Whippomorpha, excluding Hippopotamus amphibius. Within Ruminantia, eight species were domesticated and 74 species were wild.

To assemble a reference of neuropeptide prohormone proteins, predicted sequences for each genome were obtained using an established bioinformatics workflow [15, 16]. Because many Cetartiodactyla genomes lack formal annotation, supplementary steps were added to detect uncharacterized prohormone loci. The longest protein isoforms were collected from a curated set of 98 mammalian prohormone sequences previously reported for this clade [15, 17, 18, 25–27]. Potential gene locations were searched using TBLASTN [28] with default parameters but with the low-complexity filter turned off. For each hit, roughly 10,000 bp of surrounding genomic sequence was extracted, and proteins were predicted from that segment using Genewise [29] with standard settings. Manual curation confirmed that predicted sequences matched the longest known isoforms. Variations in genome completeness occasionally produced truncated or duplicated sequences; in those cases, further analyses were conducted with modified parameters, broader intronic coverage, and relaxed composition-based filters [30] to refine predictions.

Once the neuropeptide gene sequences were validated, interspecific comparisons were performed. Protein alignments were generated for each prohormone using MAFFT [31], applying the L-INS-i algorithm suited for domains surrounded by variable regions [31]. Phylogenetic reconstructions of each gene were produced with PhyML [32], and the resulting trees were visualized through ASTRAL software [33–36]. The Pearson correlation was used to evaluate the concordance between species trees and individual gene trees, and mean protein evolutionary distance values were calculated to quantify divergence both within and among taxonomic groups [37].

Finally, the evolutionary and domestication framework best supported by data from all 118 species was determined. This analysis incorporated parameters related to amino acid aromaticity (A), composition (C), polarity (P), and side-chain volume (V) [38–40]. Substitution rates were modeled according to the Jones–Taylor–Thornton (JTT) matrix [41]. The most appropriate model for the dataset was identified using the Akaike Information Criterion (AIC) to select the configuration with the strongest statistical support [42].

Results and Discussion

Identification of prohormones among species

Across the 118 Cetartiodactyla genomes examined, all 98 neuropeptide prohormone proteins were detected, though their distribution varied between taxonomic groups. One major observation was the absence of relaxin 1 (RLN1) in Ruminantia, whereas it remained detectable in all other suborders of Cetartiodactyla. Conversely, peptide YY2 (PYY2) appeared exclusively within Ruminantia—fully identifiable in Giraffidae and Bovinae and partially in Cervidae. The proteins encoded by AVP, secretin (SCT), tachykinin precursor 4 (TAC4), and VGF were partially or completely retrieved in about 20% of the examined species. Meanwhile, apelin (APLN), CART pre-propeptide (CARTPT), NPY, and somatostatin (SST) displayed very high conservation across taxa.

At the species level, genome completeness strongly influenced gene recovery. For example, argali (Ovis ammon), Alpine ibex (Ammotragus lervia), and blue whale (Balaenoptera musculus) permitted detection of roughly 56%, 53%, and 79% of neuropeptide prohormones, respectively. In the hartebeest (Alcelaphus buselaphus), around 69% were partially or fully identified. Variability in sequencing depth, assembly continuity, or annotation quality—rather than biological divergence—largely explained incomplete sequence reconstruction. Assembly gaps, frame

inconsistencies, and splice site prediction errors constrained accurate evaluation of true prohormone gene gain or loss events, despite adjustments to the prediction workflow.

Species tree based on prohormone sequences

Phylogenetic reconstruction

The reconstructed species tree aligned well with the expected evolutionary structure of the four Cetartiodactyla suborders **(Figure 1)**. Because RLN1 and PYY2 were limited to specific suborders, they were excluded from this tree. Within Whippomorpha, the hippopotamus appeared as a more distant lineage from Cetacea, which itself split into the parvorders Mysticeti (baleen whales) and Odontoceti (toothed whales). The phylogeny also confirmed the consistent grouping of Ruminantia members into their recognized families and subfamilies.

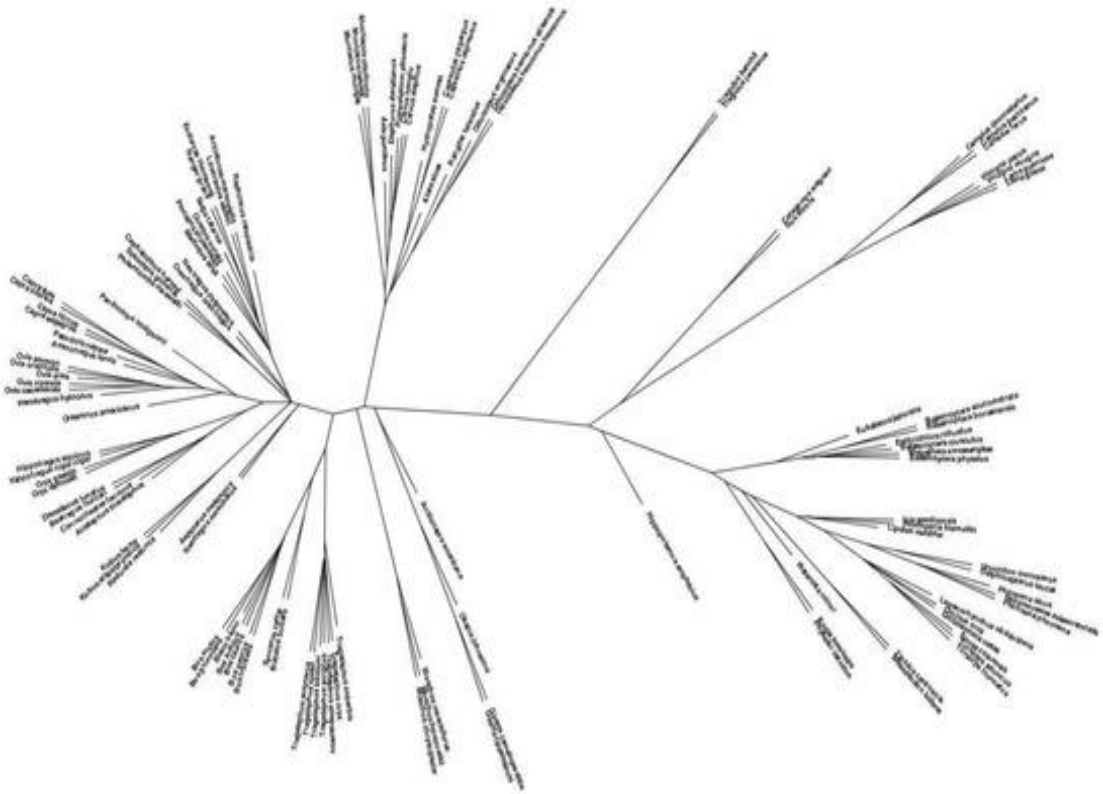

Figure 1. Species tree reconstructed from individual prohormone gene trees.

Correlation between Interspecies distances derived from neuropeptide genes

To evaluate divergence consistency, pairwise species distances derived from individual prohormone trees were compared to those from the overall species tree. The mean correlation between the two datasets was 0.77, ranging from 0.25 to 0.92. For most genes (69 of 98), correlations were high (0.8–0.9). Highly conserved peptides tended to yield lower correlations than variable ones because they exhibited minimal sequence divergence. Genes such as chromogranin A (CHGA), prodynorphin (PDYN), and thyrotropin-releasing hormone (TRH) achieved correlations above 0.9, while AVP, insulin (INS), natriuretic peptide C (NPPC), prokineticin 2 (PROK2), parathyroid hormone (PTH), and SST showed weaker correspondence (< 0.50) due to strong conservation.

Within Ruminantia, correlations between gene-level and species-level distances were comparatively higher, likely reflecting the dominance of ruminant representatives in the dataset. In contrast, the Whippomorpha suborder—particularly the hippopotamus, sperm whale (Physeter catodon), and baiji (Lipotes vexillifer)—displayed greater interspecies divergence.

Evolutionary model analysis

An evolutionary model incorporating amino acid aromaticity (A), composition (C), polarity (P), and side-chain volume (V) was applied to examine sequence adaptation across the 118 species and suborders. **Table 2** lists the

neuropeptide genes according to parameter behavior. Most genes had at least one nonzero parameter estimate: A and V often exhibited negative coefficients, whereas C and P generally showed positive ones. These patterns paralleled correlations among amino acid properties—A and V were positively correlated (~ 0.7) but inversely related to C and P (− 0.35 to − 0.47). Parameters C and P were moderately correlated (0.4), which strengthened to 0.81 when cysteine residues were omitted.

Table 2. Distribution of neuropeptide prohormone genes according to parameter direction in the evolutionary model.

Taxonomic Group[1]	mS[2]	Parameter[3]	A	C	P	V	−	+	−	+	−	+	−	+
Overall														
All	81.5		52	5	4	64	17	19	51	3				
Domestic	10		31	5	2	32	11	13	26	5				
Wild	75.5		50	6	3	62	17	18	49	6				
Wild terrestrial	56		51	3	1	56	11	23	48	3				
Ruminantia														
All	55		51	5	3	50	13	24	51	5				
Domestic	6		23	5	3	27	5	12	18	4				
Wild	52		43	5	0	54	11	22	47	4				
Bovidae	38		35	5	6	44	10	17	38	5				
Bovidae Antilopinae	11		18	8	2	24	7	4	24	3				
Bovidae Bovinae	10		22	1	6	20	10	6	15	6				
Bovidae Caprinae	7		13	7	4	22	8	14	25	2				
Cervidae	11		11	2	6	2	7	1	4	0				
Tylopoda														
All	4		20	1	13	10	9	7	10	1				
Whippomorpha														
All	20		32	3	6	23	13	11	39	0				
Cetacea Mysticeti	5		16	6	4	16	18	8	13	2				
Cetacea Odontoceti	14		30	3	7	21	11	10	31	3				

1 Taxonomic level: "All" = all species combined or grouped by suborder.

2 Median number of species per gene.

3 Number of genes with negative (−) or positive (+) estimates for each parameter (A, C, P, V).

Each gene could display multiple modified parameters

The domestic species group exhibited a greater proportion of neuropeptide prohormone genes with nonzero parameter values (40%) compared to their wild counterparts, particularly for the C parameter. In contrast, differences in parameter estimates between wild and domestic Ruminantia were relatively limited. Seven neuropeptide prohormone genes displayed parameter modifications restricted to either the domestic (1 gene) or the wild terrestrial (6 genes) categories. These genes included *apelin* (APLN), *adenylate cyclase activating polypeptide 1* (ADCYAP1), *arginine vasopressin* (AVP), *cholecystokinin* (CCK), *growth hormone releasing hormone* (GHRH), *torsin family two member A* (salusin-related isoform, TOR2X), and *VGF nerve growth factor inducible* (VGF). **Table 3** outlines the average magnitude of parameter variation for these genes. A positive estimate for parameter A in APLN suggested that domesticated species show increased representation of aromatic amino acids. Conversely, wild groups exhibited negative A and V parameters but positive C estimates, indicating a reduced preference for aromatic residues in the context of domestication.

Table 3. Parameter estimates for neuropeptide prohormone genes displaying shifts in domestic or wild species.

Parameter[1]	Symbol[2]	A		C		P		V	
		All	Rum	All	Rum	All	Rum	All	Rum
Domestic	APLN	3.62	9.80	0.00	0.00	0.00	0.00	0.00	0.00

Wild terrestrial								
	ADCYAP1	−1.04	−1.27	0.00	0.00	0.00	−0.95	0.00	0.00
	AVP	0.00	0.00	0.74	1.22	0.00	0.00	0.00	0.00
	CCK	0.00	1.70	2.03	2.18	0.00	0.00	0.00	−0.90
	GHRH	−1.35	−1.26	0.00	0.00	0.00	0.00	0.00	0.00
	TOR2X	0.00	0.00	0.00	0.00	0.00	0.00	−1.32	−1.57
	VGF	0.00	0.00	0.00	0.00	0.00	0.00	−1.23	−1.39

1 Estimated changes for aromaticity (A), composition (C), polarity (P), and volume (V) within all domestic or wild taxa (All) and within domestic or wild Ruminantia (Rum).

2 Gene abbreviations: APLN = apelin; ADCYAP1 = adenylate cyclase activating polypeptide 1; AVP = arginine vasopressin; CCK = cholecystokinin; GHRH = growth hormone releasing hormone; TOR2X = torsin family 2 member A (salusin-containing isoform); VGF = VGF nerve growth factor inducible.

Prohormone complement

The combined use of bioinformatic prediction, sequence compilation, and comparative analysis provided valuable insight into neuropeptide gene changes associated with both evolutionary diversification and domestication across Cetartiodactyla lineages. As expected from such a large and diverse order, genome completeness and sequencing quality varied considerably among assemblies. The resulting Cetartiodactyla phylogeny was nearly identical to the established taxonomic relationships [6, 43, 44]. Employing closely related reference species during annotation [45] allowed recovery of otherwise fragmented sequences and yielded consistent alignments across families, reducing potential bias introduced by incomplete assemblies.

A particularly interesting observation was that *Suina* species retained the AUG start codon for *neuropeptide W* (NPW), while all other Cetartiodactyla examined utilized a non-AUG initiation site, consistent with several other mammalian taxa [46]. The loss of the canonical AUG codon may lead to inaccurate start-site predictions for NPW. This difference is notable since pigs—representing Suina—are frequently used as biomedical analogs due to greater genomic similarity to humans than to rodents, and NPW influences numerous behavioral and physiological pathways.

Another result from the computational pipeline involved sequence variation within *cortistatin* (CORT). A 15-nucleotide insertion was detected in the signal peptide region of domestic goat CORT, yet it does not alter the mature cortistatin peptides. This insertion, present in assemblies from four distinct domestic goat breeds, was absent from other *Capra* or *Caprinae* species, suggesting a recent lineage-specific mutation.

Comparative sequence evaluation of predicted prohormones refined the understanding of *relaxin 1* (RLN1), *peptide YY2* (PYY2), *islet amyloid polypeptide* (IAPP), and *galanin-like peptide* (GALP). The current study confirmed the loss of RLN1 previously noted in cattle [17] to encompass all Ruminantia following divergence from Whippomorpha. Conversely, PYY2 was exclusively present among Ruminantia. Other suborders showed incomplete or ambiguous predictions—particularly within Whippomorpha—supporting prior conclusions that PYY2 functions as a pseudogene beyond Ruminantia [47]. GALP was broadly found across species, though terminal regions differed among suborders, and several Bovidae lacked the canonical start segment. Meanwhile, both Tylopoda and Cetacean IAPP sequences contained premature stop codons and absent signal peptides, consistent with pseudogene status. All examined camelids and the Chacoan peccary (*Catagonus wagneri*), consistent with previous porcine findings [18], lacked normal cleavage sites, although transcriptomic data confirmed IAPP expression in pigs [48].

Cross-species comparisons further revealed differences in genes potentially affecting levels of active neuropeptides, including *tachykinin precursor 4* (TAC4), *insulin-like 6* (INSL6), *gonadotropin-releasing hormone 2* (GNRH2), and *prolactin-releasing hormone* (PRLH). TAC4 sequences varied among Cetartiodactyla, but none included the cleavage motif reported in human or mouse homologs [49]. Similarly, the start and terminal sequences of INSL6 in pigs and camelids differed from other Cetartiodactyla but resembled human and rodent orthologs [50]. While the N-terminal cleavage site of the INSL6 A-chain was conserved, the C-terminal sites for both chains remain undefined in these species. For GNRH2, all Bovinae exhibited a six–amino acid extension in the signal peptide, whereas Whippomorpha lacked the expected gonadoliberin-2 motif. Additionally, all Cetacean species carried an eight–amino acid deletion within the PRLH signal peptide—absent in *Hippopotamus amphibius* and other taxa—resulting in a truncated prolactin-releasing peptide 31 (PrRP31). Conversely, a 10-base-pair

insertion shared by all Ruminantia created a longer PRLH terminal segment, though that region does not correspond to any characterized active peptide.

The variability in POMC protein architecture among Cetartiodactyla members appears to influence the production of several neuropeptides. Although a full-length POMC sequence was predicted for Tylopoda species, this lineage lacks the N-terminal peptide (NPP) cleavage motif required for synthesizing pro-γ-MSH and its derivative, the amidated γ1-MSH. No peptides originating from the pro-γ-MSH section were found through mass spectrometry [27], whereas γ1-MSH has been identified in humans [51] and cattle [52]. In contrast, every other examined Cetartiodactyla species retained this N-terminal cleavage signal. Comparable findings have been reported in Rodentia [53], where Muridae species are missing the same motif, unlike other rodent families. Notably, both Tylopoda and Muridae encode the longer γ3-MSH peptide. γ-MSH peptides are functionally associated with sodium control and blood pressure regulation [54-56]; additionally, administering γ1-MSH to the left ventral tegmental area of rats provokes grooming activity [57].

Peptides encoded by calcitonin (CALC) gene families exhibit multiple physiological functions, encompassing calcium balance, vasodilation, inflammatory response, pain regulation, migraine, and thermoregulation [58-61]. In relation to behavioral domestication, a single-nucleotide polymorphism (SNP) located within the calcitonin receptor-stimulating peptide (CRSP) gene cluster distinguishes purebred dogs—those bred under human selection—from unregulated, freely mating dogs [62]. Furthermore, expression of the calcitonin receptor-like receptor (CALCRL) differed between tame and aggressive fox phenotypes in the pituitary gland [63]. The detection of four calcitonin or CRSP genes across all Cetartiodactyla lineages in the current dataset matches the pattern observed in most mammals, with the exception of primates and rodents [64]. Although calcitonin-related genes were present, analysis of their proteolytic cleavage regions revealed missing sites necessary for generating certain peptides, such as calcitonin gene-related peptide 1, in both CRSP2 and CRSP3 sequences.

Evolutionary model

Variation across the amino acid aromaticity (A), composition (C), polarity (P), and side-chain volume (V) parameters in the evolutionary model allowed recognition of trends in amino acid property shifts in prohormone and derived neuropeptide sequences among Cetartiodactyla taxa. The comparable magnitude of changes observed between the total species set and the Ruminantia group largely reflected the overrepresentation of Ruminantia, whereas smaller clades or subfamilies displayed greater sequence divergence. Within Tylopoda, the elevated rate of modification likely results from the small number of species and their close phylogenetic relationships. Within higher taxonomic levels, Whippomorpha demonstrated a prominent share of neuropeptide prohormone genes exhibiting A, C, and V shifts, similar to the patterns seen among Mysticeti and Odontoceti families.

Most neuropeptide prohormone genes differed among Cetartiodactyla suborders; however, approximately 18%—including PMCH, GHRH, PROK2, AVP, PDGFB, and PPY—remained unchanged in the studied parameters. The preservation of these sequences aligns with amino acid replacement rates predicted by the JTT model.

Among genes that did show parameter alterations, the majority displayed modifications in either one (31%) or two (34%) parameters. Reductions predominated in A and V, while increases were most frequent in C. Genes such as HAMP, INSL3, MLB, and SPX showed these shifts across nearly all taxonomic divisions.

Analysis of coefficient patterns suggests that prohormone sequence evolution tends to limit bulky and hydrophobic residues (phenylalanine, tryptophan, and tyrosine), which could negatively influence neuropeptide function. Selective pressure thus favors smaller or less hydrophobic amino acids. Moreover, the cation-π interactions formed by aromatic side chains—non-covalent forces important for protein folding and molecular binding with neurotransmitters and drugs like serotonin [65]—appear generally disadvantageous in prohormone regions.

Domestication

The examination of neuropeptide gene sequence alterations linked to domestication incorporated the spatial distribution of both wild and domesticated taxa. Phylogenetic inference was separated orthogonally from taxonomic constraints to minimize confounding influences, particularly differences in species number and restricted interbreeding. The evolutionary framework revealed divergence between domestic and wild terrestrial categories. Domesticated Ruminantia exhibited a marginally greater proportion of sequence modifications than their wild counterparts, with approximately 60% of neuropeptide prohormone genes showing either identical

parameters or at least one overlapping value between domestic and wild groups. These outcomes imply that part of the variation in parameters originates from lineage distinctions rather than direct domestication effects.

Interpreting the relationship between domestication and neuropeptide evolution is further complicated by inconsistencies in domestication status among species and by uneven genome assembly quality. For instance, the gayal (Bos frontalis) is frequently regarded as a domesticated derivative of the gaur (Bos gaurus) [66], though domestication intensity varies substantially within the gayal population. Similarly, identical nucleotide and protein sequences obtained from all camelid genomes analyzed here contradict the earlier proposal of adaptive introgression involving endothelin 3 (EDN3) in South American camelids [6]. Moreover, ancestral hybridization events and artificial selection pressures often obscure clear associations between domestication and prohormone sequence divergence [3, 4, 6, 67]. After accounting for taxonomic distribution, levels of domestication, and differences in assembly quality, this study identified consistent amino acid parameter alterations in several neuropeptide prohormone genes linked with calm, herd-oriented behavioral traits.

The neuropeptide prohormone genes exhibiting parameter variation between domestic and wild groups encode peptides that perform diverse biological roles. Neuropeptides derived from APLN, AVP, TOR2X (torsin family two member A, salusin isoform), and VGF regulate angiogenesis and vascular tone through vasodilatory and vasoconstrictive pathways. Similarly, peptides originating from ADCYAP1, CCK, GHRH, and VGF contribute to appetite regulation and energy balance. VGF also participates in the modulation of circadian timing, nociception, learning, and memory processes [68, 69]. Within the scope of domestication, these physiological roles converge on the sympatho-adrenomedullary axis—the system governing “fight or flight” responses—where elevations in blood pressure and glucose occur during stress adaptation [70].

Conclusion

This investigation analyzed the relationship between evolutionary pressures, domestication, and changes in neuropeptide gene sequences. A dedicated bioinformatics workflow was implemented to determine the prohormone repertoire comprising 98 distinct sequences from 118 Cetartiodactyla species, both wild and domesticated, across four suborders. Comprehensive comparisons of these prohormone sequences, coupled with evolutionary modeling, revealed alterations in amino acid features such as aromaticity, composition, polarity, and molecular volume. Noteworthy findings included disruptions in cleavage motifs within specific genes (e.g., INSL6, GNRH2, PRLH, POMC, CALC) that may hinder proper neuropeptide formation. In addition, coefficient estimates indicated that evolutionary dynamics tend to select against bulky, hydrophobic residues within prohormone sequences. Modeling outcomes also suggested that genes associated with the sympatho-adrenomedullary network and the “fight or flight” mechanism could be influenced by domestication. Collectively, the prohormone dataset established here provides a structural and comparative foundation for future neuropeptidomic investigations in Cetartiodactyla, particularly those addressing traits of medical or agricultural significance.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Zeller U, Göttert T. The relations between evolution and domestication reconsidered—Implications for systematics, ecology, and nature conservation. Glob Ecol Conserv. 2019;20:e00756.
2. Zeder MA. Out of the Fertile Crescent: The dispersal of domestic livestock through Europe and Africa. In: Petraglia M, Boivin N, Crassard R, editors. Human Dispersal and Species Movement: From Prehistory to the Present. Cambridge: Cambridge University Press; 2017. p. 261-303.
3. McHugo GP, Dover MJ, MacHugh DE. Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. BMC Biol. 2019;17:98.

4. Frantz LAF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. Nat Rev Genet. 2020;21:449-60.
5. Lott R, Williamson C. Cetaceans in Captivity. In: Butterworth A, editor. Marine Mammal Welfare: Human Induced Change in the Marine Environment and its Impacts on Marine Mammal Welfare. Cham: Springer International Publishing; 2017. p. 161-81.
6. Fan R, Gu Z, Guang X, Marin JC, Varas V, Gonzalez BA, Wheeler JC, Hu Y, Li E, Sun X, et al. Genomic analysis of the domestication and post-Spanish conquest evolution of the llama and alpaca. Genome Biol. 2020;21:159.
7. Roed KH, Flagstad O, Nieminen M, Holand O, Dwyer MJ, Rov N, Vila C. Genetic analyses reveal independent domestication origins of Eurasian reindeer. Proc Biol Sci. 2008;275:1849-55.
8. Herbeck YE, Gulevich RG. Neuropeptides as facilitators of domestication. Cell Tissue Res. 2019;375:295-307.
9. Irwin DM. Variation in the Evolution and Sequences of Proglucagon and the Receptors for Proglucagon-Derived Peptides in Mammals. Front Endocrinol. 2021;12:700066.
10. Wolf Horrell EM, Boulanger MC, D'Orazio JA. Melanocortin 1 Receptor: Structure, Function, and Regulation. Front Genet. 2016;7:95.
11. de Simoni Gouveia JJ, da Silva MV, Paiva SR, de Oliveira SM. Identification of selection signatures in livestock species. Genet Mol Biol. 2014;37:330-42.
12. Dib L, San-Jose LM, Ducrest AL, Salamin N, Roulin A. Selection on the Major Color Gene Melanocortin-1-Receptor Shaped the Evolution of the Melanocortin System Genes. Int J Mol Sci. 2017;18:2618.
13. Von Heijne G. The signal peptide. J Membr Biol. 1990;115:195-201.
14. Southey BR, Rodriguez-Zas SL, Sweedler JV. Prediction of neuropeptide prohormone cleavages with application to RFamides. Peptides. 2006;27:1087-98.
15. Southey BR, Sweedler JV, Rodriguez-Zas SL. A python analytical pipeline to identify prohormone precursors and predict prohormone cleavage sites. Front Neuroinform. 2008;2:7.
16. Southey BR, Romanova EV, Rodriguez-Zas SL, Sweedler JV. Bioinformatics for Prohormone and Neuropeptide Discovery. Methods Mol Biol. 2018;1719:71-96.
17. Southey BR, Rodriguez-Zas SL, Sweedler JV. Characterization of the prohormone complement in cattle using genomic libraries and cleavage prediction approaches. BMC Genom. 2009;10:228.
18. Porter KI, Southey BR, Sweedler JV, Rodriguez-Zas SL. First survey and functional annotation of prohormone and convertase genes in the pig. BMC Genom. 2012;13:582.
19. Hu CK, Southey BR, Romanova EV, Maruska KP, Sweedler JV, Fernald RD. Identification of prohormones and pituitary neuropeptides in the African cichlid, Astatotilapia burtoni. BMC Genom. 2016;17:660.
20. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L. The origin of the domestic pig: Independent domestication and subsequent introgression. Genetics. 2000;154:1785-91.
21. Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, Maisano Delser P, Hare AJ, Burger J, Collins MJ, Kehati R, et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science. 2019;365:173-6.
22. Ciani E, Mastrangelo S, Da Silva A, Marroni F, Ferencakovic M, Ajmone-Marsan P, Baird H, Barbato M, Colli L, Delvento C, et al. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genet Sel Evol. 2020;52:25.
23. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, Funk K, Ketter A, Kim S, Kimchi A, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2020;48:D9-16.
24. The Bactrian Camels Genome Sequencing and Analysis Consortium, Jirimutu, Wang Z, Ding G, Chen G, Sun Y, Sun Z, Zhang H, Wang L, Hasi S, et al. Genome sequences of wild and domestic bactrian camels. Nat Commun. 2012;3:1202.
25. Amare A, Hummon AB, Southey BR, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV. Bridging neuropeptidomics and genomics with bioinformatics: Prediction of mammalian neuropeptide prohormone processing. J Proteome Res. 2006;5:1162-7.
26. Tegge AN, Southey BR, Sweedler JV, Rodriguez-Zas SL. Comparative analysis of neuropeptide cleavage sites in human, mouse, rat, and cattle. Mamm Genome. 2008;19:106-20.

27. Alim FZD, Romanova EV, Tay YL, Rahman A, Chan KG, Hong KW, Rogers M, Southey BR, Greenwood MP, Mecawi AS, et al. Seasonal adaptations of the hypothalamo-neurohypophyseal system of the dromedary camel. PLoS ONE. 2019;14:e0216679.
28. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-402.
29. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14:988-95.
30. Yu YK, Wootton JC, Altschul SF. The compositional adjustment of amino acid substitution matrices. Proc Natl Acad Sci USA. 2003;100:15688-93.
31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772-80.
32. Gil M, Zanetti MS, Zoller S, Anisimova M. CodonPhyML: Fast maximum likelihood phylogeny estimation under codon substitution models. Mol Biol Evol. 2013;30:1270-80.
33. Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. ASTRAL: Genome-scale coalescent-based species tree estimation. Bioinformatics. 2014;30:i541-8.
34. Mirarab S, Warnow T. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 2015;31:i44-52.
35. Yin J, Zhang C, Mirarab S. ASTRAL-MP: Scaling ASTRAL to very large datasets using randomization and parallelization. Bioinformatics. 2019;35:3961-9.
36. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform. 2018;19:153.
37. Wise MJ. Mean protein evolutionary distance: A method for comparative protein evolution and its application. PLoS ONE. 2013;8:e61276.
38. Sneath PH. Relations between chemical structure and biological activity in peptides. J Biol. 1966;12:157-95.
39. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185:862-4.
40. Xia X, Li WH. What amino acid properties affect protein evolution? J Mol Evol. 1998;47:557-64.
41. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8:275-82.
42. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19:716-23.
43. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364:eaav6202.
44. Heckeberg NS. The systematics of the Cervidae: A total evidence approach. PeerJ. 2020;8:e8114.
45. Southey BR, Rodriguez-Zas SL, Rhodes JS, Sweedler JV. Characterization of the prohormone complement in Amphiprion and related fish species integrating genome and transcriptome assemblies. PLoS ONE. 2020;15:e0228562.
46. Tanaka H, Yoshida T, Miyamoto N, Motoike T, Kurosu H, Shibata K, Yamanaka A, Williams SC, Richardson JA, Tsujino N, et al. Characterization of a family of endogenous neuropeptide ligands for the G protein-coupled receptors GPR7 and GPR8. Proc Natl Acad Sci USA. 2003;100:6251-6.
47. Couzens M, Liu M, Tuchler C, Kofler B, Nessler-Menardi C, Parker RM, Klocker H, Herzog H. Peptide YY-2 (PYY2) and pancreatic polypeptide-2 (PPY2): Species-specific evolution of novel members of the neuropeptide Y gene family. Genomics. 2000;64:318-23.
48. Southey BR, Zhang P, Keever MR, Rymut HE, Johnson RW, Sweedler JV, Rodriguez-Zas SL. Effects of maternal immune activation in porcine transcript isoforms of neuropeptide and receptor genes. J Integr Neurosci. 2021;20:21-31.
49. Zhang Y, Lu L, Furlonger C, Wu GE, Paige CJ. Hemokinin is a hematopoietic-specific tachykinin that regulates B lymphopoiesis. Nat Immunol. 2000;1:392-7.
50. Lok S, Johnston DS, Conklin D, Lofton-Day CE, Adams RL, Jelmberg AC, Whitmore TE, Schrader S, Griswold MD, Jaspers SR. Identification of INSL6, a new member of the insulin family that is expressed in the testis of the human and rat. Biol Reprod. 2000;62:1593-9.
51. Fenger M, Johnsen AH. Alpha-amidated peptides derived from pro-opiomelanocortin in normal human pituitary. Biochem J. 1988;250:781-8.
52. Bohlen P, Esch F, Shibasaki T, Baird A, Ling N, Guillemin R. Isolation and characterization of a gamma 1-melanotropin-like peptide from bovine neurointermediate pituitary. FEBS Lett. 1981;128:67-70.

53. Jégou S, Cone RD, Eberlé AN, Vaudry H. Chapter 111—Melanocortins. In: Kastin AJ, editor. Handbook of Biologically Active Peptides. 2nd ed. Boston, MA: Academic Press; 2013. p. 838-44.
54. Wikberg JE, Muceniece R, Mandrika I, Prusis P, Lindblom J, Post C, Skottner A. New aspects on the melanocortins and their receptors. Pharm Res. 2000;42:393-420.
55. Humphreys MH. Gamma-MSH, sodium metabolism, and salt-sensitive hypertension. Am J Physiol Regul Integr Comp Physiol. 2004;286:R417-30.
56. Harmer SC, Bicknell AB. Role of gamma-MSH peptides in the regulation of adrenal steroidogenesis. Peptides. 2005;26:1944-51.
57. Klusa V, Svirskis S, Opmane B, Muceniece R, Wikberg JE. Behavioural responses of gamma-MSH peptides administered into the rat ventral tegmental area. Acta Physiol Scand. 1999;167:99-104.
58. Nyberg F, Hallberg M. Neuropeptides in hyperthermia. Prog Brain Res. 2007;162:277-93.
59. Russell FA, King R, Smillie SJ, Kodji X, Brain SD. Calcitonin gene-related peptide: Physiology and pathophysiology. Physiol Rev. 2014;94:1099-142.
60. Granstein RD, Wagner JA, Stohl LL, Ding W. Calcitonin gene-related peptide: Key regulator of cutaneous immunity. Acta Physiol. 2015;213:586-94.
61. Brain SD, Grant AD. Vascular actions of calcitonin gene-related peptide and adrenomedullin. Physiol Rev. 2004;84:903-34.
62. Pilot M, Malewski T, Moura AE, Grzybowski T, Olenski K, Kaminski S, Fadel FR, Alagaili AN, Mohammed OB, Bogdanowicz W. Diversifying Selection Between Pure-Breed and Free-Breeding Dogs Inferred from Genome-Wide SNP Analysis. G3. 2016;6:2285-98.
63. Hekman JP, Johnson JL, Edwards W, Vladimirova AV, Gulevich RG, Ford AL, Kharlamova AV, Herbeck Y, Acland GM, Raetzman LT, et al. Anterior Pituitary Transcriptome Suggests Differences in ACTH Release in Tame and Aggressive Foxes. G3. 2018;8:859-73.
64. Katafuchi T, Yasue H, Osaki T, Minamino N. Calcitonin receptor-stimulating peptide: Its evolutionary and functional relationship with calcitonin/calcitonin gene-related peptide based on gene structure. Peptides. 2009;30:1753-62.
65. Dougherty DA. Cation-pi interactions involving aromatic amino acids. J Nutr. 2007;137:1504S-8S.
66. Ahrestani FS. Bos frontalis and Bos gaurus (Artiodactyla: Bovidae). Mamm Species. 2018;50:34-50.
67. Frantz LA, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, Paudel Y, Crooijmans RP, Larson G, Groenen MA. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet. 2015;47:1141-8.
68. Lewis JE, Brameld JM, Jethwa PH. Neuroendocrine Role for VGF. Front Endocrinol. 2015;6:3.
69. Chen S, Reichert S, Singh C, Oikonomou G, Rihel J, Prober DA. Light-Dependent Regulation of Sleep and Wake States by Prokineticin 2 in Zebrafish. Neuron. 2017;95:153-68.e156.
70. Sapolsky RM, Romero LM, Munck AU. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr Rev. 2000;21:55-89.