



Corticotropin-Releasing Factor: An Ancient Peptide Family Related to the Secretin Peptide Superfamily

Ola M. Michalec¹, Belinda S. W. Chang¹, Nathan R. Lovejoy² and David A. Lovejoy^{1*}

¹ Department of Cell and Systems Biology, University of Toronto - St. George Campus, Toronto, ON, Canada, ² Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada

OPEN ACCESS

Edited by:

Vance L. Trudeau,
University of Ottawa, Canada

Reviewed by:

Richard Giuseppe Manzon,
University of Regina, Canada

James A. Carr,
Texas Tech University, United States

*Correspondence:

David A. Lovejoy
david.lovejoy@utoronto.ca

Specialty section:

This article was submitted to
Neuroendocrine Science,
a section of the journal
Frontiers in Endocrinology

Received: 20 April 2020

Accepted: 29 June 2020

Published: 27 August 2020

Citation:

Michalec OM, Chang BSW,
Lovejoy NR and Lovejoy DA (2020)
Corticotropin-Releasing Factor: An
Ancient Peptide Family Related to the
Secretin Peptide Superfamily.
Front. Endocrinol. 11:529.
doi: 10.3389/fendo.2020.00529

Corticotropin-releasing factor (CRF) is the hypothalamic releasing peptide that regulates the hypothalamic-pituitary-adrenal/inter-renal (HPA/I) axis in vertebrates. Over the last 25 years, there has been considerable discussion on its paralogous genes, urotensin-I/urocortin-1, and urocortins-2 and-3 and their subsequent role in the vertebrate stress response. Phylogenetically, the CRF family of peptides also belong to the diverse assemblage of Secretin- and Calcitonin-based peptides as evidenced by comparative-based studies of both their ligand and G-protein-coupled receptor (GPCR) structures. Despite this, the common origin of this large assemblage of peptides has not been ascertained. An unusual peptide, teneurin-C-terminal associated peptide (TCAP), reported in 2004, comprises the distal extracellular tip of the teneurin transmembrane proteins. Further studies indicated that this teneurin region binds to the latrophilin family of GPCRs. Initially thought to be a member of the Secretin GPCR family, evidence indicates that the latrophilins are a member of the Adhesion family of GPCRs and are related to the common ancestor of both Adhesion and Secretin GPCR families. In this study, we posit that TCAP may be a distantly related ancestor of the CRF-Calcitonin-Secretin peptide family and evolved near the base of metazoan phylogeny.

Keywords: secretin superfamily, TCAP, teneurin, adhesion GPCRs, evolution, metabolism

INTRODUCTION

Corticotropin-releasing factor (CRF) is the critical hypothalamic releasing factor that regulates the hypothalamus-pituitary-adrenal/inter-renal (HPA/1) axis in vertebrates, yet after some 40 years after its discovery, numerous questions still exist regarding when, why, and how this peptide evolved. We hypothesize that due to the high level of primary structure similarity among CRF paralogs and related peptide lineages (e.g., calcitonin, secretin) there was likely an ancestor peptide common to this cluster. We further suggest that the “teneurin C-terminal associated peptides” (TCAP) represent an extant candidate lineage related to the hypothetical common ancestor.

The discovery of CRF in the early 1980s (1) occurred about the same time as the discovery of other peptides of similar structure [sauvagine (2); urotensin-I (3)]. Later, Vale and his laboratory characterized a mammalian version of sauvagine/urotensin-I in rat brain that they termed, urocortin (4). Further phylogenetic studies suggested that mammalian urocortin, amphibian sauvagine, and fish urotensin-I were orthologs of the same gene (5). In 2001, the structures of two novel related peptides were reported by the Vale laboratory who named the peptides, urocortin 2

and 3 (6, 7) and by Hsu and Hsieh (8) who termed the peptides as “stresscopin” and “stresscopin-related peptide.” These novel CRF family homologs were subsequently established to be a separate paralog lineage of CRF and the urotensin-I/sauvagine/urocortin grouping (5, 9–13). In parallel to studies of vertebrate CRF isoforms, the presence of related peptides were reported in insects and arthropods (12, 14–16). Therefore, the high degree of structural similarity among CRF-like peptides in both deuterostome (e.g., chordates) and protostomes (e.g., arthropods) indicated that an ancestral peptide with CRF family primary structure attributes was present before the bifurcation of these metazoan lineages. Importantly, this ancestral peptide appeared to exist in a physiologically mature form indicative of a distant lineage that likely radiated as other ancestral peptides with distinct but overlapping functions. The identity of these hypothetical ancestral peptides has remained elusive, however, it is plausible that these lineages led to the evolution and expansion of the secretin and calcitonin family of peptides (11, 12).

The Secretin superfamily of peptides is a diverse assemblage of peptide lineages with overlapping functions utilizing structurally related receptors. The nomenclature describing the phylogeny of the secretin grouping of peptides and receptors is confusing. In order to clarify this, we have used the term “secretin family” to denote those peptides that are thought to be part of a direct monophyletic clade (e.g., secretin, PACAP, VIP, and glucagon paralogs). For the inclusion of the wider group which include CRF and calcitonin, we have referred to this as the “Secretin superfamily.” Due, in part, to the similarity and structural conservation of their cognate receptors, the Secretin family G-Protein Coupled Receptors (GPCR) was defined as a distinct clade (17). The Secretin superfamily of peptides is one of the five main families of ligands that bind to GPCRs. The GPCRs have most recently been classified into five main families using the GRAFS system; Glutamate (G), Rhodopsin (R), Adhesion (A), Frizzled/Taste2 (F), and Secretin (S) (17). Notably, both CRF and calcitonin receptors are included within the Secretin GPCR family. Among these, Adhesion and Secretin GPCRs are the most evolutionarily ancient (18). Adhesion GPCRs have a characteristically long N-terminus rich in serine and threonine residues whereas Secretin GPCRs have a characteristic hormone-binding domain (HBD) in their N-terminal region (18). Secretin-related receptors form a single monophyletic clade that derived from the Adhesion GPCRs (18, 19). Adhesion GPCR genes have been identified in choanoflagellate and sea anemone genomes but Secretin GPCR genes have not suggesting that Adhesion GPCRs are more evolutionarily ancient than Secretin GPCRs (18). Interestingly, some derived phylogenetically younger Adhesion GPCR members possess an HBD with highly conserved amino acid sequences and similar splice site motifs as Secretin GPCRs. These observations led, in part, to the hypothesis that the Secretin GPCR clade was derived from an offshoot of the Adhesion GPCR lineage. However, although the data linking the Adhesion and Secretin superfamilies were compelling, evidence of a structurally related peptide ligand linking the two receptor clades was lacking.

One such lineage of Adhesion GPCRs that does possess a HBD with similar structural motifs to Secretin GPCRs

are the latrophilins (LPHN) or ADGRL (Adhesion G-protein coupled receptors, subfamily L). It was originally considered a new type of Secretin GPCR, due to its characteristic HBD, but has now been re-classified as an Adhesion GPCR (17). The first identified ligand for ADGRL was α -latrotoxin, a peptide component of black widow spider toxin venom that specifically targets vertebrates (20) and shares major sequence similarity with other Secretin superfamily ligands (21). The data suggest that these peptides have a common origin. Although, α -latrotoxin was an exogenous ligand, the high affinity binding of this soluble peptide to ADGRL indicated that this receptor had the potential to bind and be activated by an endogenous peptide similar to the structure of α -latrotoxin. The search for this theoretical ligand led to the identification of the teneurin transmembrane proteins as a likely suspect.

Several recent studies established that the distal region of the extracellular domain of the teneurin transmembrane proteins binds ADGRL with high affinity and activates the receptor. Silva et al. (22) first discovered that teneurin-2, expressed on post-synaptic dendritic branches, binds LPHN-1 expressed on pre-synaptic nerve terminals to form a trans-synaptic complex. Similar trans-cellular interactions were observed between teneurins-2 and 4 and all three LPHNs (23) and between teneurin-1 and LPHN-3 (24). A C-terminal fragment of teneurin-2, named Lasso, triggered an increase in cytosolic Ca^{2+} in Nb2a cells overexpressing LPHN-1 and in pre-synaptic nerve terminals of hippocampal cells (22). This distal region of the teneurin extracellular domain contains a peptide-like sequence termed “teneurin C-terminal associated peptide” (TCAP). The TCAPs are a family of four bioactive peptides that are 40–41 amino acids in length and are located at the C-terminus of each of the teneurin transmembrane proteins (25, 26). TCAPs possess a cleavage motif at the N-terminus and an amidation motif at the C-terminus (27) and may be autolytically cleaved from teneurins upon binding with LPHN (28, 29). TCAP shares about 20% sequence similarity with CRF and calcitonin, members of the Secretin superfamily of ligands, suggesting a common evolutionary origin (27). Moreover, our laboratory has recently identified that teneurin C-terminal associated peptide (TCAP)-1 is likely an endogenous ligand that interacts with the HBD of LPHN (30).

Therefore, as TCAP binds to an Adhesion GPCR and shares sequence similarity to CRF and calcitonin, ligands that bind to Secretin GPCR receptors that are classified as being most closely related to ancestral Adhesion GPCRs, this prompted the investigation of TCAP as a progenitor of the Secretin superfamily. The hypothesis that the teneurin-TCAP system is an ancient system that arose prior to the emergence of the Metazoa as a result of a horizontal gene transfer (HGT) event from a prokaryote to a choanoflagellate ancestor has previously been raised (31–33). However, the TCAP family has not been previously examined. Thus, TCAP may be associated with an early evolving lineage of peptides that is a sister lineage to the CRF, calcitonin, and secretin families of peptides (11, 34). We therefore examined the phylogenetic relationships of these peptides using TCAP as an outgroup.

MATERIALS AND METHODS

Collection of Sequences

Peptide sequences of Secretin GPCR ligands, including CRF, calcitonin and secretin families, and Adhesion GPCR ligands, including TCAP 1–4, as well as reference groups including neuropeptide Y (NPY) and insulin were collected among a range of extant protostomes and deuterostomes, using the GenBank genome sequence analysis program on the NCBI website. The peptides were organized by organism, phylum, class, and order and were tabulated and their accession numbers were recorded (**Table 1**). Sequences were divided into pre-propeptides (or propeptides for TCAP) and mature peptides, after which were imported to MEGA 6.0 for analysis (38). Downloaded from <http://www.megasoftware.net/>.

Sequence Alignments

Peptide sequences were aligned using the MUSCLE algorithm (39). The alignment was examined, reviewed for duplicate sequences using pairwise distances ($d = 0.0$ was identical) and excess sequence was cut at both 5' and 3' ends, as these fragments did not contribute to the alignment. Modifications to the alignment were made to ensure that the characteristic residue motifs were conserved. This included highly conserved cysteine (C), tryptophan (W), arginine (R), and lysine (K) residues throughout as well as motifs characteristic of each family. For the CRF family this was the 5' leucine (L), serine (S), and the 3' asparagine (N) motif that is conserved throughout the entire family, the "TCV" or "TCXV" motif that is conserved among the calcitonin family and the "PELAD" motif that is conserved among the TCAP family.

Phylogenetic Analysis

Phylogenetic tree construction and statistical analyses were carried out in MEGA 6.0 (38). A multi-step approach was undertaken in order to understand the relationship of each family relative to TCAP prior to conducting a comprehensive analysis of all of the families.

Maximum Likelihood (ML) Method

The amino acid substitution model and the rate among sites were both chosen based on the model that resulted in the greatest log likelihood, the lowest Akaike Information Criterion (AIC) and the lowest Bayesian Information Criterion (BIC), parameters calculated by MEGA 6.0. To ensure the most accurate analysis, these parameters were calculated for each constructed tree. The model that maximized the log likelihood was used for analysis. A partial deletion of sequences with too many gaps/missing data was applied with a cutoff of 95%, so sites that were not found in at least 95% of sequences were not used toward the analysis. The applied heuristic method was Nearest-Neighbor Interchange (NNI), so the initial trees were obtained using the NJ method to a matrix of pairwise distances estimated using a JTT model. Reliability of the tree was tested using 1,000 bootstrap replicates.

Pre-propeptide and Mature Peptide Analysis

Two sets of analyses were performed. The first involved Secretin superfamily pre-propeptides, which are composed of

a signal, cryptic, and mature peptide and TCAP propeptides, as TCAP does not possess a signal peptide. Given the functional importance, bioactivity, and high level of conservation throughout evolution, a second separate analysis was performed on mature peptides of both Secretin superfamily and TCAP family members.

For analysis involving Secretin superfamily pre-propeptides and TCAP family propeptides, a total of 181 amino acid sequences were used, with a total of 44 positions in the final dataset after all positions with <95% site coverage were eliminated.

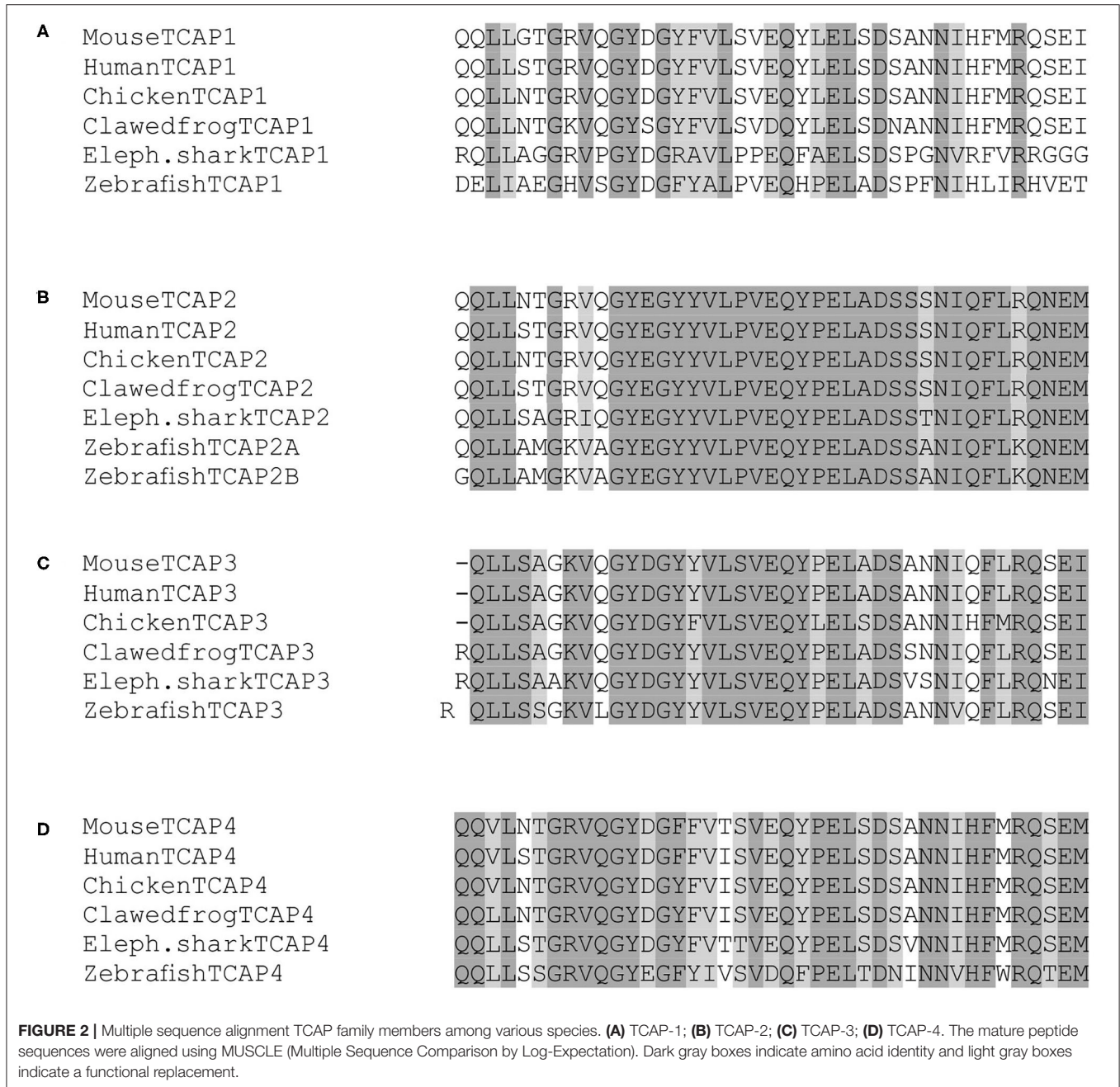
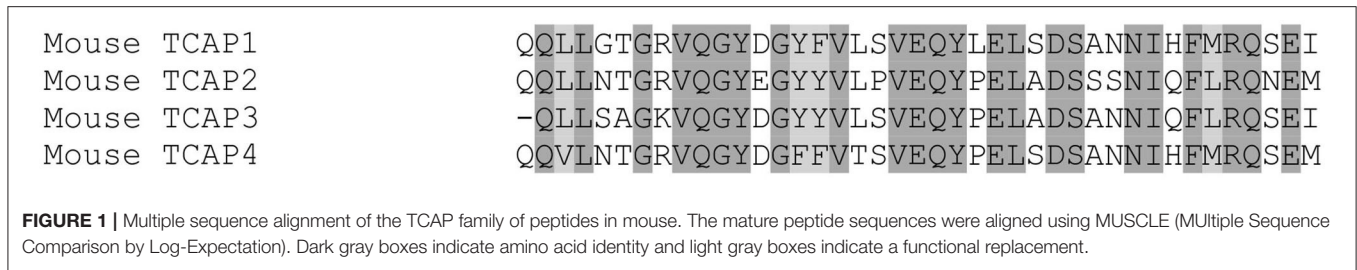
Mature Peptide Analysis

For analysis involving Secretin superfamily mature peptides and TCAP mature peptides, a multi-step analysis was undertaken in order to elucidate the relationships of each family with respect to one another and TCAP. As insulin has a tertiary structure where the peptide folds and the two mature chains are connected by sets of disulfide bonds from the cysteine residues (40), the mature peptide had to be divided into A and B chains for the purpose of this analysis. Due to the high sequence conservation of NPY that may have resulted in the odd placement of the NPY reference group in the pre-propeptide analysis and given that the NPY mature peptide is even so more highly conserved, it was not included as a reference group in the analysis of mature peptides. For insulin and the calcitonin family, analysis involved 72 amino acid sequences, with a total of 14 positions in the final dataset. For insulin, calcitonin, and TCAP, analysis involved 95 amino acid sequences, with a total of 14 positions in the data set. For insulin, calcitonin, CRF and TCAP families, analysis involved 135 amino acid sequences leaving 12 positions in the final data set. Lastly, for insulin, calcitonin, CRF, secretin, and TCAP families, analysis included 179 amino acid sequences leaving 15 positions in the final dataset.

RESULTS

Sequence Analysis of TCAP Paralogs and Orthologs

TCAP paralogs, those that diverged as a result of a genome duplication event, demonstrated a high degree of conservation (**Figure 1**). When TCAP 1–4 are aligned in mouse, residues Q2, L4, G7, V9, Q10, G11, Y12, G14, V17, V20, E21, Q22, Y23, E25, L26, D28, S29, N32, I33, F35, R37, Q38, and E40 are all conserved among the four paralogs (**Figure 1**). Similarly, TCAP orthologs, those that arose as a result of a species divergence, demonstrate a high degree of conservation among vertebrates (**Figure 2**). When mammalian, bird, amphibian, and fish TCAP 1–4 sequences are aligned residues L3, G7, V9, G11, Y12, G14, L18, Q22, E25, L26, D28, N32, R37 are conserved among TCAP-1 orthologs (**Figure 2A**). Among TCAP-2 orthologs, residues Q2, L3, L4, G7, G11, Y12, E13, G14, Y15, Y16, V17, L18, P19, V20, E21, Q22, Y23, P24, E25, L26, A27, D28, S29, S30, N32, I33, Q34, F35, L36, Q38, N39, E40, M41 are conserved (**Figure 2B**). Among TCAP-3 orthologs, Q2, L3, L4, S5, K8, V9, G11, Y12, D13, G14, Y15, V17, L18, S19, V20, E21, Q22, Y23, E25, L26, D28, S29, N32, F35, R37, Q38, E40, I41 are conserved (**Figure 2C**). Lastly, among TCAP-4



orthologs, Q1, Q2, L4, G7, R8, V9, Q10, G11, Y12, G14, F15, V20, Q22, P24, E25, L26, D28, N31, N32, H34, F35, R37, Q38, E40,

M41. Overall, TCAP-2 orthologs (**Figure 2B**) are the most highly conserved and TCAP-1 orthologs (**Figure 2A**) are the least highly

conserved. Also, a characteristic “PELAD” motif at positions 24–28 from the N-terminus is conserved among the TCAP paralogs and orthologs. The high level of conservation of the “PELAD” motif suggests that it possesses a functional attribute, such as a receptor binding or activation site (27). This family of peptides contains the “PELAD” motif at residues 24–28 from the N-terminus. Therefore, both TCAP orthologs and paralogs demonstrate a high degree of conservation among vertebrates.

Evolutionary Analysis of Pre-propeptides and Mature Peptides of Secretin Superfamily and TCAP Family Members

Phylogenetic analysis of CRF, calcitonin, and secretin pre-propeptide families and TCAP family propeptides revealed that each family formed a distinct group. TCAP, CRF, and secretin families form distinct clades and insulin forms a sister group with the calcitonin family (Figure 3). Also, CRF and calcitonin are closely related sister lineages and they, in turn, form a sister lineage to the secretin family. TCAP, the putative progenitor, is most distantly related to these families relative to their relationships to one another.

A separate analysis was performed with mature peptide sequences of the Secretin superfamily and TCAP mature peptides due to their high conservation and functional importance throughout evolution. Phylogenetic analysis of calcitonin mature peptides, insulin A and B mature chains and TCAP demonstrated that calcitonin and insulin families are sister lineages (Figure 4). Insulin A chains were more closely related to the calcitonin family than insulin B chains (Figure 4). Phylogenetic analysis of calcitonin, insulin A and B chains, CRF, and TCAP mature peptides confirmed that calcitonin and insulin families were sister lineages and that CRF formed a separate group to these two families (Figure 5). Lastly, phylogenetic analysis of calcitonin, insulin A and B chains, CRF, secretin, and TCAP mature peptides revealed that calcitonin and insulin families were sister lineages and that both CRF and secretin formed separate groups from these two families (Figure 6). Therefore, the multi-step mature peptide analysis confirmed that insulin and calcitonin are sister lineages, that form distinct groups from CRF and secretin families and in turn, that the TCAP family is a distinct clade from Secretin superfamily members.

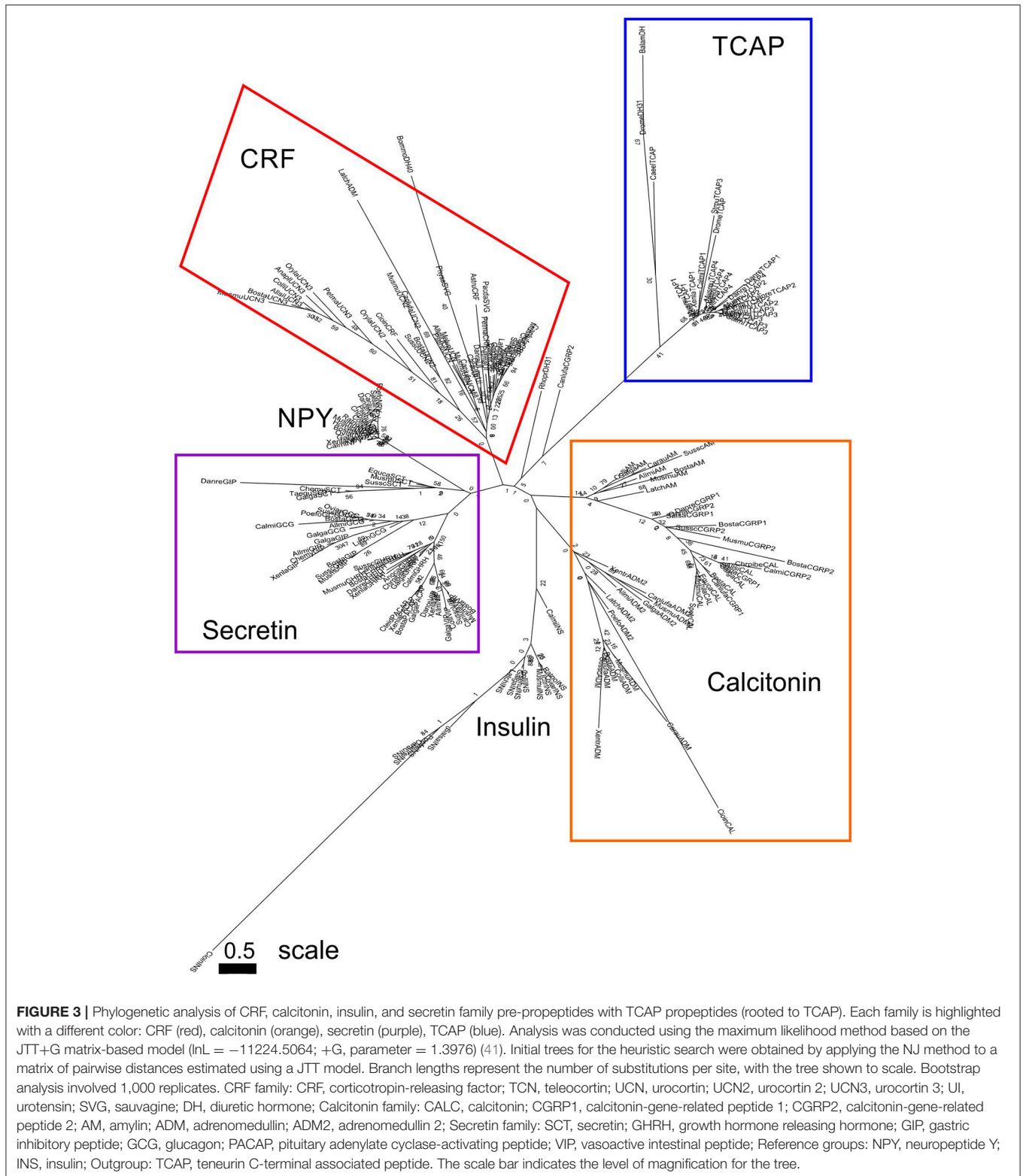
DISCUSSION

In this study, the TCAP family is presented as a putative progenitor of the Secretin superfamily of ligands for the first time. The evolutionary relationships among the receptors of these peptides are well-established (18, 19). However, the relationships among members of the Secretin superfamily of ligands as well as a progenitor for this family of peptides have not been elucidated. We considered TCAP as a putative progenitor of the Secretin superfamily for the following reasons. First, evolutionary relationships among the receptors of these ligands demonstrate that Secretin GPCRs derived from Adhesion GPCRs (19) and as TCAP-1 binds to LPHN, an Adhesion GPCR with a HBD characteristic of Secretin GPCRs (17). It is possible that a

similar course of evolution occurred for the ligands. Second, the sequence similarity that TCAP shares with CRF and calcitonin (27), both Secretin superfamily members whose receptors are the most closely related to Adhesion GPCRs, suggests that these peptides may have evolved from TCAP, a candidate progenitor peptide.

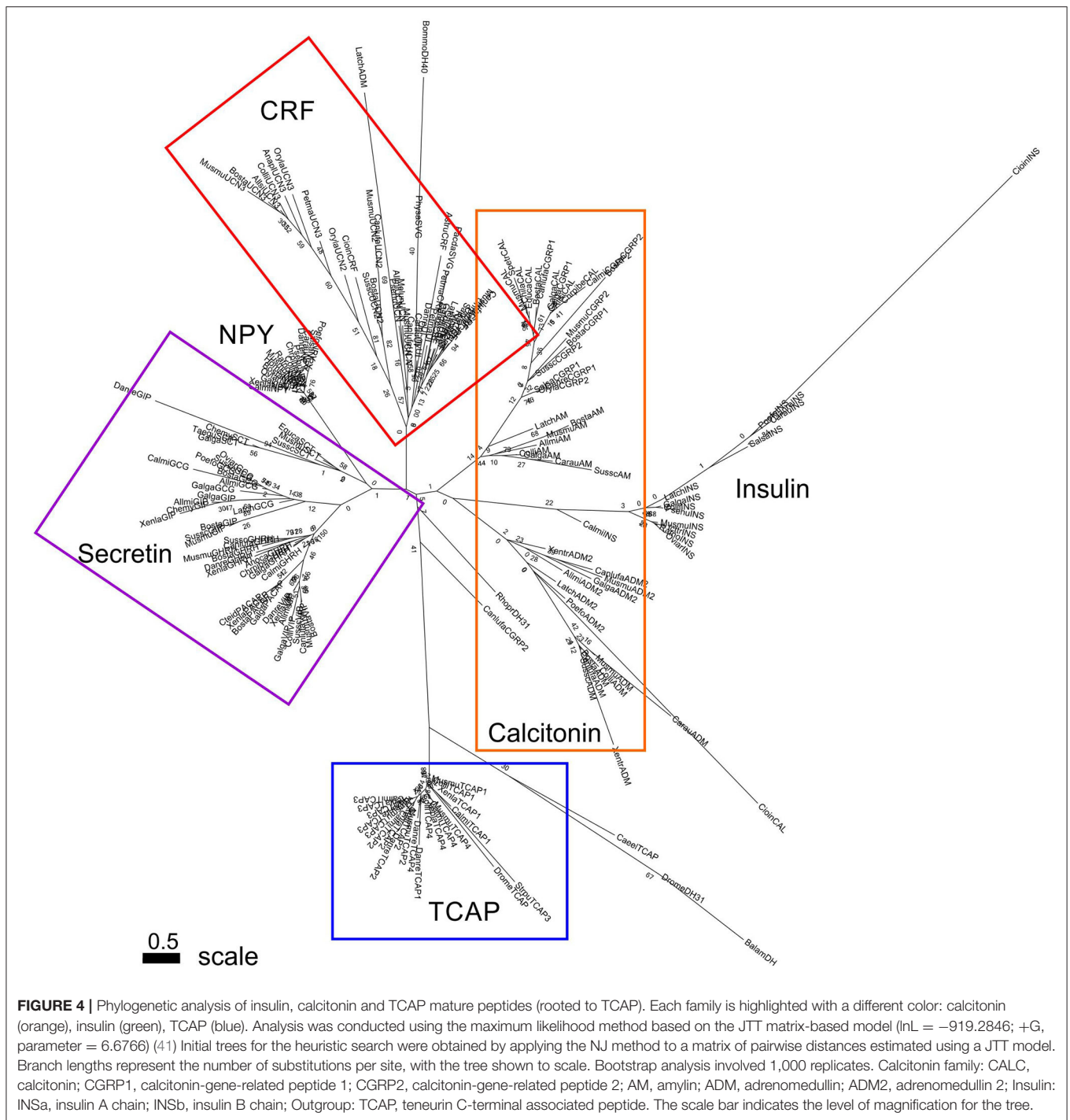
The teneurin-TCAP system is well-established as being evolutionarily ancient. Evidence suggests that this system arose before the Metazoa evolved about 1 billion years ago and prior to the emergence of the Secretin superfamily that arose around the time of the protostome-deuterostome divergence, about 600 million years ago. As a result, although the TCAP sequence shows some amino acid similarity with the Secretin superfamily, there are a number of differences indicating that the two lineages are evolutionarily divergent. Indeed, we could not determine any significant binding or activation capacity of TCAP with any members of the Secretin GPCRs [(11, 34); Lovejoy, unpublished observations]. In contrast, TCAP binds to the latrophilin HBD and activates this receptor [(30); Reid et al., submitted]. As proposed by Zhang et al. (33), the teneurin-TCAP system likely evolved from a polymorphic proteinaceous toxin (PPT) gene that arose as a result of a HGT event from a prokaryote to a choanoflagellate, a primitive unicellular organism. Importantly, the teneurin gene has been identified in the choanoflagellate, *Monosiga brevicollis* (32). Choanoflagellates are thought to be a progenitor to the Metazoans (42). This supports the hypothesis that a choanoflagellate may have engulfed a prokaryote containing the PPT gene, which became integrated into its genome and lost its toxic role over time (32, 33). With respect to structural evidence, the teneurins share characteristics of PPTs: the same type II orientation, rearrangement hotspot (RHS) domains and close similarity to the C-terminal domain to the histidine-asparagine-histidine (HNH) bacterial toxin of the glycine-histidine-histidine (GHH) clade (33, 43). The GHH domain may be an ancestor of TCAP that lost its toxic role and functioned as an intracellular signaling molecule (33). Additionally, the C-terminal region of the *M. brevicollis* teneurin protein contains tyrosine-aspartate (YD) repeats characteristic of proteobacteria and most of the extracellular domain is encoded on one large 6,829 base pair exon characteristic of prokaryotic genomes and of HGT (32). Therefore, evidence suggests that the teneurin-TCAP system is ancient as it evolved as a result of a HGT event prior to the emergence of the Metazoa.

Moreover, with respect to the course of evolution of the receptors, evidence demonstrates that Adhesion GPCRs evolved prior to Secretin GPCRs and that Secretin GPCRs are derived from Adhesion GPCRs. Adhesion GPCR genes have been identified in the genome of amphioxus, *Branchiostoma floridae*, the choanoflagellate, *M. brevicollis*, and the sea anemone, *Nematostella vectensis* (18), meaning that these lineages were present prior to the protostome-deuterostome divergence. On the other hand, Secretin GPCRs have not been identified in these species and therefore, receptor lineages of the Secretin superfamily likely expanded and radiated around the time of the bifurcation of protostomes and deuterostomes. Also, Nordström et al. (18) demonstrated Secretin GPCRs evolved from Adhesion



GPCRs using phylogenetic analysis. Therefore, evidence that the teneurin-TCAP system arose prior to the emergence of the Metazoa as well as the characterization of Adhesion GPCRs

but not Secretin GPCRs prior to the protostome-deuterostome divergence suggests that the teneurin-TCAP system predates members of the Secretin superfamily. We suggest that if the



ligands for these receptors underwent a similar course in evolution, the TCAP family may be a putative progenitor to the Secretin superfamily.

In light of the evidence to suggest that the teneurin-TCAP system evolved prior to the emergence of the Metazoa, the previously established relationship that Secretin GPCRs derived from Adhesion GPCRs [(Nordstrom et al., 2009); (19)], the evidence that TCAP binds to LPHN, an Adhesion

GPCR with a HBD characteristic of Secretin GPCRs (17) and given the sequence similarity that TCAP shares with Secretin superfamily members, CRF, and calcitonin (27), a phylogenetic investigation using TCAP as a putative progenitor of the Secretin superfamily was undertaken. A putative progenitor of the Secretin superfamily of ligands has not been previously established. Sequence analysis of TCAP family members demonstrated a highly conserved peptide and

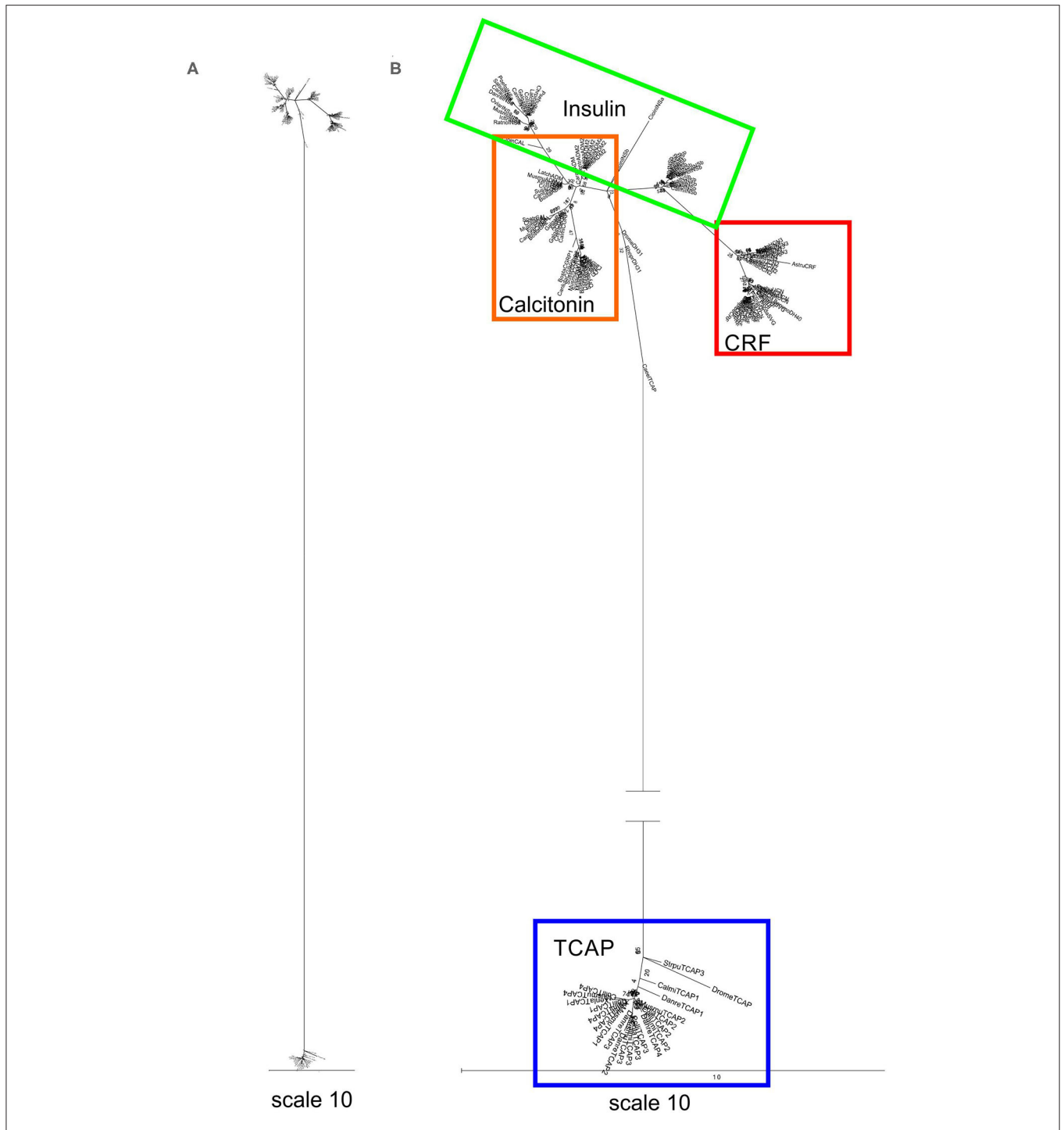


FIGURE 5 | Phylogenetic analysis of insulin, calcitonin, CRF and TCAP mature peptides (rooted to TCAP). The trees are represented as **(A)** original tree with the appropriate scale **(B)** magnified and rooted to TCAP. Each family is highlighted with a different color: CRF (red), calcitonin (orange), insulin (green), TCAP (blue). Analysis was conducted using the maximum likelihood method based on the Dayhoff matrix-based model (lnL = -1019.5552; +G, parameter = 6.6766) (41). Initial trees for the heuristic search were obtained by applying the NJ method to a matrix of pairwise distances estimated using a JTT model. Branch lengths represent the number of substitutions per site, with the tree shown to scale. Bootstrap analysis involved 1,000 replicates. Calcitonin family: CALC, calcitonin; CGRP1, calcitonin-gene-related peptide 1; CGRP2, calcitonin-gene-related peptide 2; AM, amylin; ADM, adrenomedullin; ADM2, adrenomedullin 2; Insulin: INSa, insulin A chain; INSc, insulin B chain; CRF family: CRF, corticotropin-releasing factor; TCN, teleocortin; UCN, urocortin; UCN2, urocortin 2; UCN3, urocortin 3; UI, urotensin; SVG, sauvagine; DH, diuretic hormone; Outgroup: TCAP, teneurin C-terminal associated peptide. The scale bar indicates the level of magnification for the tree.

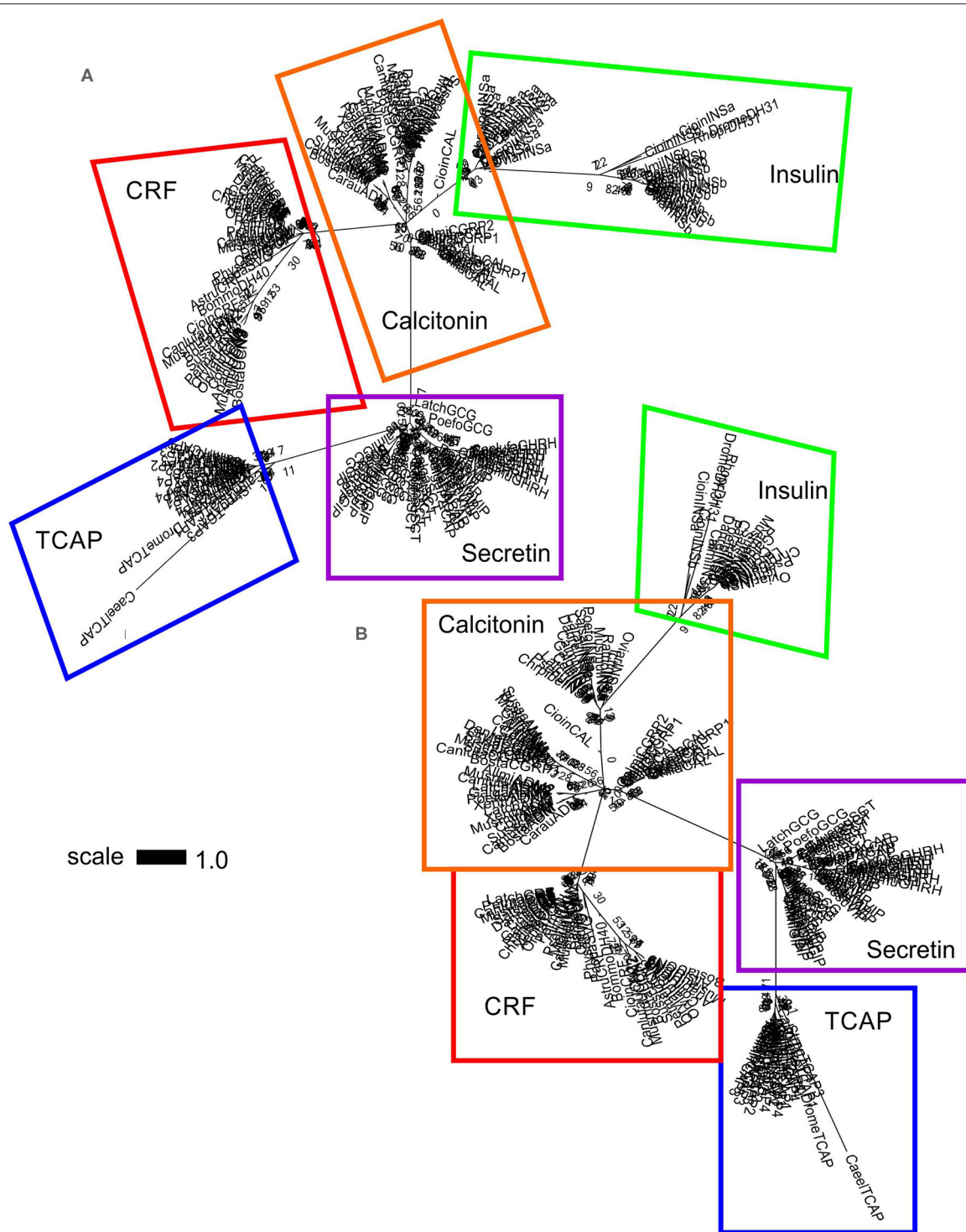


FIGURE 6 | Phylogenetic analysis of insulin, calcitonin, CRF, secretin, and TCAP mature peptides. The trees are represented as **(A)** unrooted and **(B)** rooted to TCAP. Each family is highlighted with a different color: CRF (red), calcitonin (orange), insulin (green), secretin (purple), and TCAP (blue). Analysis was conducted using the maximum likelihood method based on the Whelan and Goldman model (lnL = -1781.0007; +G, parameter = 23.5912) (41). Initial trees for the heuristic search were obtained by applying the NJ method to a matrix of pairwise distances estimated using a JTT model. Branch lengths represent the number of substitutions per site, with the tree shown to scale. Bootstrap analysis involved 1,000 replicates. Calcitonin family: CALC, calcitonin; CGRP1, calcitonin-gene-related peptide 1; CGRP2, calcitonin-gene-related peptide 2; AM, amylin; ADM, adrenomedullin; ADM2, adrenomedullin 2; INSA, insulin A chain; INSB, insulin B chain; CRF family: CRF, corticotropin-releasing factor; TCN, teleocortin; UCN, urocortin; UCN2, urocortin 2; UCN3, urocortin 3; UI, urotensin; SVG, sauvagine; DH, diuretic hormone; Secretin family: SCT, secretin; GHRH, growth hormone releasing hormone; GIP, gastric inhibitory peptide; GCG, glucagon; PACAP, pituitary adenylate cyclase-activating peptide; VIP, vasoactive intestinal peptide; Outgroup: TCAP, teneurin C-terminal associated peptide. The scale bar indicates the level of magnification for the tree.

TABLE 1 | List of Sequences used to construct phylogenetic trees.

Hormone	Organism	Phylum	Class	Order	Accession number
CRF FAMILY					
CRF	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	NM_001172210
CRF	<i>Rana sylvatica</i> (Wood frog)	Chordata	Amphibia	Anura	HQ630608
CRF	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_001123031
CRF	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	XM_005506466
CRF	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	JN859047
CRF	<i>Salmo salar</i> (Atlantic salmon)	Chordata	Actinopterygii	Salmoniformes	NM_001141590
CRF	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_205769 XM_357335
CRF	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	NM_001014278 XM_544106
CRF	<i>Petromyzon marinus</i> (Sea lamprey)	Chordata	Cephalaspidomorphi	Petromyzontiformes	Endsin et al. (35)
CRF	<i>Latimeria chalumnae</i> (Coelacanth)	Chordata	Sarcopterygii	Coelacanthiformes	XM_006009123
CRF	<i>Chrysemys picta bellii</i> (Western painted turtle)	Chordata	Reptilia	Testudines	XM_005288085
CRF	<i>Asterias rubens</i> (Common starfish)	Echinodermata	Asteroidea	Forcipulatida	Semmens et al. (36)
CRF	<i>Ciona intestinalis</i> (Sea tunicate)	Urochordata	Ascidiacea	Enterogona	Lovejoy and Barsyte-Lovejoy, (37)
TCN	<i>Oryzias latipes</i> (Japanese rice fish)	Chordata	Actinopterygii	Beloniformes	BAR90710.1
SVG	<i>Pachymedusa dancicolor</i> (Mexican tree frog)	Chordata	Amphibia	Anura	FR846380
SVG	<i>Phyllomedusa sauvagii</i>	Chordata	Amphibia	Anura	AAY21509.1
UCN	<i>Pseudopodoces humilis</i> (Tibetan ground-tit)	Chordata	Aves	Passeriformes	XM_005532743
UCN	<i>Melopsittacus undulatus</i> (Budgerigar)	Chordata	Aves	Psittaciformes	XM_005140144
UCN	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_021290
UCN	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	XM_848667
UCN	<i>Alligator mississippiensis</i> (American alligator)	Chordata	Reptilia	Crocodylia	XM_006273603
UCN2	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	XM_003132185
UCN2	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	XM_005223016
UCN2	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_145077 XM_910603
UCN2	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	XM_005632587
UCN2	<i>Oryzias latipes</i> (Japanese rice fish)	Chordata	Actinopterygii	Beloniformes	NP_001121991.1
UCN3	<i>Anas platyrhynchos</i> (Mallard)	Chordata	Aves	Anseriformes	XM_005014551
UCN3	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_031250
UCN3	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_001076527
UCN3	<i>Alligator sinensis</i> (Chinese alligator)	Chordata	Reptilia	Crocodylia	XM_006035307
UCN3	<i>Petromyzon marinus</i> (Sea lamprey)	Chordata	Cephalaspidomorphi	Petromyzontiformes	Endsin et al. (35)
UCN3	<i>Oryzias latipes</i> (Japanese rice fish)	Chordata	Actinopterygii	Beloniformes	NP_001121992.1
UI	<i>Oncorhynchus mykiss</i> (Rainbow trout)	Chordata	Actinopterygii	Salmoniformes	NM_001124343
UI	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NM_001030180 XM_687090
UI	<i>Carassius auratus</i> (Goldfish)	Chordata	Actinopterygii	Cypriniformes	AF129115
UI	<i>Petromyzon marinus</i> (Sea lamprey)	Chordata	Cephalaspidomorphi	Petromyzontiformes	Endsin et al. (35)
DH 31	<i>Drosophila melanogaster</i> (Fruit fly)	Arthropoda	Insecta	Diptera	NM_078790
DH 40	<i>Bombyx mori</i> (Domestic silkworm)	Arthropoda	Insecta	Lepidoptera	AB298934
DH 31	<i>Rhodnius prolixus</i> (Assassin bug)	Arthropoda	Insecta	Hemiptera	HM030716
DH	<i>Balanus amphitrite</i> (Barnacle)	Arthropoda	Maxillopoda	Sessilia	JQ864196
CALCITONIN FAMILY					
CALC	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	X03012
CALC	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	AB462435
CALC	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	X97991
CALC	<i>Alligator sinensis</i> (Chinese alligator)	Chordata	Reptilia	Crocodylia	XM_006018232
CALC	<i>Chrysemys picta bellii</i> (Western painted turtle)	Chordata	Reptilia	Testudines	XM_005303304
CALC	<i>Ciona intestinalis</i> (Sea tunicate)	Urochordata	Ascidiacea	Enterogona	AB485672
CALC	<i>Equus caballus</i> (Horse)	Chordata	Mammalia	Perissodactyla	AF249307
CALC	<i>Chinchilla lanigera</i> (Long-tailed chinchilla)	Chordata	Mammalia	Rodentia	XP_005380394.1

(Continued)

TABLE 1 | Continued

Hormone	Organism	Phylum	Class	Order	Accession number
CALC	<i>Spermophilus tridecemlineatus</i> (Squirrel)	Chordata	Mammalia	Rodentia	XM_005326775
CGRP1	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_001113708
CGRP1	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NM_001002471
CGRP1	<i>Salmo salar</i> (Atlantic salmon)	Chordata	Actinopterygii	Salmoniformes	NM_001146580
CGRP1	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_001076340
CGRP1	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	NM_001003266
CGRP2	<i>Callorhynchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XM_007887527
CGRP2	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_001134662
CGRP2	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	NM_001002948
CGRP2	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_054084
CGRP2	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	NM_001102473
CGRP2	<i>Oryzias latipes</i> (Japanese rice fish)	Chordata	Actinopterygii	Beloniformes	NM_001104894
Amylin	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	XM_005504095
Amylin	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_205397
Amylin	<i>Carassius auratus</i> (Goldfish)	Chordata	Actinopterygii	Cypriniformes	EU000530
Amylin	<i>Latimeria chalumnae</i> (Coelacanth)	Chordata	Sarcopterygii	Coelacanthiformes	XM_005998862
Amylin	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_001195038
Amylin	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_010491
Amylin	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	XM_003126437
Amylin	<i>Alligator mississippiensis</i> (American alligator)	Chordata	Reptilia	Crocodylia	XM_006270879
ADM	<i>Xenopus tropicalis</i> (Western clawed frog)	Chordata	Amphibia	Anura	XM_002936741
ADM	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	XM_005499377
ADM	<i>Carassius auratus</i> (Goldfish)	Chordata	Actinopterygii	Cypriniformes	EU000533
ADM	<i>Latimeria chalumnae</i> (Coelacanth)	Chordata	Sarcopterygii	Coelacanthiformes	XM_005998689
ADM	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	D14875
ADM	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_173888
ADM	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	AB191461
ADM	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_009627
ADM2	<i>Xenopus tropicalis</i> (Western clawed frog)	Chordata	Amphibia	Anura	XM_002939371
ADM2	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	XM_004937395
ADM2	<i>Latimeria chalumnae</i> (Coelacanth)	Chordata	Sarcopterygii	Coelacanthiformes	XM_006013419
ADM2	<i>Poecilia formosa</i> (Amazon molly)	Chordata	Actinopterygii	Cyprinodontiformes	XM_007545759
ADM2	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	XM_843399
ADM2	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	AB121035
ADM2	<i>Alligator mississippiensis</i> (American alligator)	Chordata	Reptilia	Crocodylia	XM_006275255
SECRETIN FAMILY					
GHRH	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	NM_001096728
GHRH	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_001040464
GHRH	<i>Callorhynchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XM_007885752
GHRH	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NM_001080092
GHRH	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_178325
GHRH	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	NM_001290112
GHRH	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_010285
GHRH	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	NM_001195118
GHRH	<i>Anolis carolinensis</i> (Green anole)	Chordata	Reptilia	Squamata	XM_003225298
GHRH	<i>Chrysemys picta bellii</i> (Western painted turtle)	Chordata	Reptilia	Testudines	XM_005295322
GIP	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	NM_001097922
GIP	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_001080104
GIP	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NM_001080059
GIP	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_001166605
GIP	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_008119

(Continued)

TABLE 1 | Continued

Hormone	Organism	Phylum	Class	Order	Accession number
GIP	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	NM_001287408
GIP	<i>Alligator mississippiensis</i> (American alligator)	Chordata	Reptilia	Crocodylia	XM_006277905
GIP	<i>Chelonia mydas</i> (Green sea turtle)	Chordata	Reptilia	Testudines	XM_007061917
GCG	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_205260
GCG	<i>Callorhynchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XM_007889848
GCG	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	NM_173916
GCG	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	NM_214324
GCG	<i>Ovis aries</i> (Sheep)	Chordata	Mammalia	Cetartiodactyla	XM_004004659
GCG	<i>Alligator mississippiensis</i> (American alligator)	Chordata	Reptilia	Crocodylia	XM_006277994
GCG	<i>Poecilia formosa</i> (Amazon molly)	Chordata	Actinopterygii	Cyprinodontiformes	XM_007546594
GCG	<i>Latimeria chalumnae</i> (Coelacanth)	Chordata	Sarcopterygii	Coelacanthiformes	XM_006004345
PACAP	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	AF187877
PACAP	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	AY956323
PACAP	<i>Ctenopharyngodon idella</i> (Grass carp)	Chordata	Actinopterygii	Cypriniformes	EF592488
PACAP	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	AY924308
SCT	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_001024833
SCT	<i>Taeniopygia guttata</i> (Zebra finch)	Chordata	Aves	Passeriformes	NM_001256233
SCT	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	X73580
SCT	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	XM_003122391
SCT	<i>Equus caballus</i> (Horse)	Chordata	Mammalia	Perissodactyla	XM_003362642
SCT	<i>Chelonia mydas</i> (Green sea turtle)	Chordata	Reptilia	Testudines	XM_007060911
VIP	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	NM_001085714
VIP	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	XM_005507654
VIP	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NM_205366
VIP	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NM_001114553
VIP	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	AF503910
VIP	<i>Canis lupus familiaris</i> (Dog)	Chordata	Mammalia	Carnivora	XM_005615524
VIP	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NM_011702
VIP	<i>Sus scrofa</i> (Pig)	Chordata	Mammalia	Artiodactyla	NM_001195233
VIP	<i>Alligator mississippiensis</i> (American alligator)	Chordata	Reptilia	Crocodylia	XM_006265239
OTHER					
NPY	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	EDK98613.1
NPY	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	AAI62071.1
NPY	<i>Carassius auratus</i> (Goldfish)	Chordata	Actinopterygii	Cypriniformes	AAA49186.1
NPY	<i>Rattus norvegicus</i> (Norway rat)	Chordata	Mammalia	Rodentia	NP_036746.1
NPY	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NP_990804.1
NPY	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	AAH80115.1
NPY	<i>Ovis aries</i> (Sheep)	Chordata	Mammalia	Cetartiodactyla	NP_001009452.1
NPY	<i>Bos taurus</i> (Cattle)	Chordata	Mammalia	Artiodactyla	ACH61954.1
NPY	<i>Homo sapiens</i> (Human)	Chordata	Mammalia	Primate	NP_000896.1
NPY	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	NP_001269740.1
NPY	<i>Pseudopodoces humilis</i> (Ground tit)	Chordata	Aves	Passeriformes	XP_005518939.1
NPY	<i>Callorhynchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	ACF22970.1
NPY	<i>Chrysemys picta bellii</i> (Western painted turtle)	Chordata	Reptilia	Testudines	XP_005290923.1
INS	<i>Ciona intestinalis</i> (Sea tunicate)	Urochordata	Ascidacea	Enterogona	NP_001123204.1
INS	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	ABF48502.1
INS	<i>Rattus norvegicus</i> (Norway rat)	Chordata	Mammalia	Rodentia	AAA41439.1
INS	<i>Ictidomys tridecemlineatus</i> (Ground squirrel)	Chordata	Mammalia	Rodentia	AAK72558.1
INS	<i>Ovis aries</i> (Sheep)	Chordata	Mammalia	Cetartiodactyla	AAB60625.1
INS	<i>Gallus gallus</i> (Chicken)	Chordata	Aves	Galliformes	NP_990553.1

(Continued)

TABLE 1 | Continued

Hormone	Organism	Phylum	Class	Order	Accession number
INS	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	EMC88047.1
INS	<i>Pseudopodoces humilis</i> (Ground tit)	Chordata	Aves	Passeriformes	XP_005522396.1
INS	<i>Callorhinchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XP_007902984.1
INS	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NP_571131.1
INS	<i>Carassius auratus</i> (Goldfish)	Chordata	Actinopterygii	Cypriniformes	ALO24192.1
INS	<i>Salmo salar</i> (Atlantic salmon)	Chordata	Actinopterygii	Salmoniformes	ACI69187.1
INS	<i>Latimeria chalumnae</i> (Coelacanth)	Chordata	Sarcopterygii	Coelacanthiformes	XP_006008147.1
INS	<i>Poecilia formosa</i> (Amazon molly)	Chordata	Actinopterygii	Cyprinodontiformes	XP_016521686.1
INS	<i>Chrysemys picta bellii</i> (Western painted turtle)	Chordata	Reptilia	Testudines	XP_005312438.1
TCAP1	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	XP_017951867.1
TCAP1	<i>Callorhinchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XP_007893009.1
TCAP	<i>Caenorhabditis elegans</i> (Roundworm)	Nematoda	Secernentea	Rhabditida	NM_171175
TCAP	<i>Drosophila melanogaster</i> (Fruit fly)	Arthropoda	Insecta	Diptera	NP_001097661
TCAP1	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NP_035985.2
TCAP1	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	XP_691552.5
TCAP1	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	EMC88689.1
TCAP2	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	XP_012815129.1
TCAP2	<i>Callorhinchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XP_007900206.1
TCAP2	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NP_035986.3
TCAP2	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	XP_017208443.1
TCAP2	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	EMC78205.1
TCAP3	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	NP_001096158
TCAP3	<i>Callorhinchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XP_007894102.1
TCAP3	<i>Strongylocentrotus purpuratus</i> (Sea urchin)	Echinodermata	Echinoidea	Echinoidea	XM_001180001
TCAP3	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NP_035987.3
TCAP3	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NP_571043.1
TCAP3	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	XP_005505621.1
TCAP4	<i>Xenopus laevis</i> (African clawed frog)	Chordata	Amphibia	Anura	NP_001096158.1
TCAP4	<i>Callorhinchus milii</i> (Elephant shark)	Chordata	Chondrichthyes	Chimaeriformes	XP_007900970.1
TCAP4	<i>Mus musculus</i> (House mouse)	Chordata	Mammalia	Rodentia	NP_001297689.1
TCAP4	<i>Danio rerio</i> (Zebrafish)	Chordata	Actinopterygii	Cypriniformes	NP_571044.2
TCAP4	<i>Columba livia</i> (Rock pigeon)	Chordata	Aves	Columbiformes	XP_005500626.1

phylogenetic analysis of the Secretin superfamily in relation to TCAP as a putative progenitor revealed relationships among Secretin superfamily members. Calcitonin and insulin families are sister lineages and they are much more closely related to one another than was previously thought. Also, calcitonin and insulin are sister lineages that form distinct lineages to CRF and secretin families. Therefore, placing TCAP as an ancestor of the Secretin superfamily allowed a novel interpretation of evolutionary relationships among Secretin superfamily members.

Sequence Analysis of TCAP Paralogs and Orthologs

Sequence analysis of both TCAP paralogs and orthologs revealed that this family of peptides is highly conserved. The presence of a conserved “PELAD” motif among TCAP orthologs and paralogs, suggests that it may possess a functional attribute, such as a receptor-binding or activation site (27). Also, some characteristic amino acids are retained throughout orthologs and paralogs. Arginine (R) and lysine (K) residues are retained in some parts

of the mature peptide and they are often characteristic of the presence of cleavage sites. Glycine (G) and proline (P) are also highly conserved and these amino acids have a tendency to be retained as their secondary structure can break the α -helical structure of peptides. A peptide system with such a large amount of conservation is indicative of great functional importance that may have been selected for. Therefore, the high sequence conservation among TCAP orthologs and paralogs suggests that this peptide system is evolutionarily ancient and may have been strongly selected for throughout evolutionary time.

Evolutionary Analysis of Pre-propeptides and Mature Peptides of Secretin Superfamily and TCAP Family Members

Phylogenetic analysis of Secretin superfamily pre-propeptides (composed of the signal, cryptic, and mature peptide) and TCAP family pro-peptides (composed of the cryptic and mature peptide) was undertaken in order to elucidate the relationships among these peptides. Analysis revealed that calcitonin, CRF,

secretin, and TCAP families formed distinct groups. Despite being chosen to serve as a reference group because it binds to a tyrosine kinase receptor and not a GPCR, insulin formed a group with calcitonin, suggesting that they may be sister lineages (Figure 3). The close relationship between calcitonin and insulin has previously been explored where Wimalawansa (44) suggested that insulin and calcitonin families are closely related. This is supported by phylogenetic analysis of the pre-propeptides and suggests that insulin and calcitonin are sister lineages. When the tree was rooted to TCAP (Figure 3), to establish the assumption that TCAP is the ancestor, CRF, calcitonin, and secretin families formed distinct groups. This evolutionary analysis suggests that the secretin family forms a separate clade that is a sister to CRF and calcitonin families, which, in turn, are sisters to one another. This is consistent with what has been observed with respect to Secretin GPCR evolution, where CRF and calcitonin receptors share the greatest amount of sequence similarity among Secretin GPCRs (17). Therefore, it is possible that a similar evolutionary scheme occurred with respect to the ligands. Thus, analysis of Secretin superfamily pre-propeptides with TCAP propeptides suggests that insulin and calcitonin are closely related sister lineages, that calcitonin-insulin and CRF lineages are closely related and that calcitonin-insulin and CRF form a distinct sister lineage to the secretin family.

Subsequently, phylogenetic analysis was performed with the mature peptides of Secretin superfamily members and the TCAP family. The analysis of TCAP family mature peptide sequences with calcitonin and insulin mature sequences (Figure 4) demonstrated that insulin A chains were closely related to mature calcitonin peptides. This suggests that the insulin A mature chain is more closely related to the calcitonin family than the insulin B mature chain, which is different from what was previously suggested by Wimalawansa (44). Subsequent analyses involving CRF, calcitonin, insulin, and TCAP mature peptides (Figure 5) as well as secretin, CRF, calcitonin, insulin, and TCAP mature peptides (Figure 6) confirmed that the insulin A chain was more closely related to the calcitonin family than the insulin B chain. Taken together, insulin and calcitonin are closely related sister groups, which was also observed with the pre-propeptide analysis (Figure 3). Moreover, with respect to relationships among Secretin superfamily members, calcitonin-insulin, and CRF families are more closely related to one another than they are to secretin or TCAP, which is supported by the evolutionary scheme of their receptors, which also appear to be very closely related. Finally, secretin forms a sister lineage to a lineage that comprises both calcitonin-insulin and CRF families. This is consistent with what was observed for analysis of the pre-propeptides (Figure 3).

Considering the evidence with respect to the ancestral origin of the teneurin-TCAP system and in light of the findings presented here, it is possible to present two hypotheses for the evolutionary scheme of these peptides. The first suggests that an ancient TCAP-like peptide may have been the ancestor of the Secretin superfamily and that it evolved prior to the emergence of CRF, calcitonin, and secretin families. This is supported by the identification of TCAP in organisms prior to

the protostome-deuterostome divergence, where as members of the Secretin superfamily have not been identified this early in evolution (31, 32, 34). The possibility of a second hypothesis, suggesting that the Secretin superfamily forms a parallel lineage to extant TCAP and that these two lineages evolved from a proto-CRF-calcitonin-secretin-TCAP ancestor that was related to all of these families, cannot be discounted. Due to sequence availability, phylogenetic analysis was performed using extant Secretin superfamily and TCAP sequences. As a result, both of these hypotheses are plausible. Future analysis should be undertaken in order to further investigate whether TCAP is a progenitor of the Secretin superfamily of ligands.

CONCLUSIONS

Taken together, phylogenetic analysis of members of the Secretin superfamily using TCAP as a putative progenitor demonstrated relationships among Secretin superfamily members. First, calcitonin formed a closely related sister lineage to insulin, particularly the insulin A chain with respect to mature peptides, but this was also observed with the pre-propeptides. Also, calcitonin-insulin and CRF families are more closely related to one another than they are to secretin or TCAP, which is supported by the evolutionary scheme of their receptors. Finally, secretin forms a sister lineage to a group that comprises both calcitonin-insulin and CRF. Therefore, given evidence that the teneurin-TCAP system arose as a result of a HGT event prior to the emergence of the Metazoa, as well as the previously established structural similarity of TCAP to calcitonin and CRF, members of the Secretin superfamily, the presented phylogenetic analysis allowed for the elucidation of relationships among members of the Secretin superfamily. To conclude, this is the first time that relationships among this family of peptides were resolved and because a progenitor peptide for the Secretin superfamily has not been elucidated, we present TCAP as a candidate progenitor.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

OM performed all analyses and completed the first draft of the paper. BC and NL provided technical guidance on the construction of the phylogenetic tree. DL oversaw the research program and completed the final draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a grant from the Natural Sciences and Engineering Council of Canada (NSERC)-Canada Graduate Scholarship-Master's (CGSM) program in addition to funding from the University of Toronto for OM.

REFERENCES

- Vale W, Spiess J, Rivier C, Rivier J. Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and β -endorphin. *Science*. (1981) 213:1394–7. doi: 10.1126/science.6267699
- Monticucci PA, Neschen A, Erspamer V. Structure of sauvagine, as vasoactive peptide from the skin of a frog. *Hoppe-Seyler's Z Physiol Chem*. (1979) 360:1178.
- Lederis K, Letter A, McMaster D, Moore G, Schlesinger D. Complete amino acid sequence of urotensin I, a hypotensive and corticotropin-releasing neuropeptide from *Catostomus*. *Science*. (1982) 218:162–4. doi: 10.1126/science.6981844
- Vaughan J, Donaldson C, Bittencourt J, Perrin MH, Lewis K, Sutton S, et al. Characterization of urocortin, a novel mammalian neuropeptide related to fish urotensin-I and to CRF. *Nature*. (1995) 378:287–92. doi: 10.1038/378287a0
- Lovejoy DA, Balment RJ. Evolution and physiology of the corticotropin-releasing factor (CRF) family of neuropeptides in vertebrates. *GenComp Endocrinol*. (1999) 155:1–22. doi: 10.1006/gcen.1999.7298
- Lewis K, Li C, Perrin MH, Blount A, Kunitake K, Donaldson C, et al. Identification of urocortin III, an additional member of the corticotropin-releasing factor (CRF) family with high affinity for the CRF2 receptor. *Proc Natl Acad Sci USA*. (2001) 98:7570–5. doi: 10.1073/pnas.121165198
- Reyes TM, Lewis K, Perrin MH, Kunitake KS, Vaughan J, Arias CA, et al. Urocortin II: a member of the corticotropin releasing factor (CRF) neuropeptide family that is selectively bound by type 2 receptors. *Proc Natl Acad Sci USA*. (2001) 98:2843–8. doi: 10.1073/pnas.051626398
- Hsu SY, Hseuh AJW. Human stresscopin and stresscopin-related peptide are selective ligands for the type 2 corticotropin-releasing hormone receptor. *Nat Med*. (2001) 7:605–11. doi: 10.1038/87936
- Cardoso JC, Bergqvist CA, Félix RC, Larhammar D. Corticotropin-releasing hormone family evolution: five ancestral genes remain in some lineages. *J Mol Endocrinol*. (2016) 57:73–86. doi: 10.1530/JME-16-0051
- Lovejoy DA. Structural evolution of urotensin-I: retaining ancestral functions before corticotropin-releasing hormone evolution. *Gen Comp Endocrinol*. (2009) 164:15–9. doi: 10.1016/j.ygcen.2009.04.014
- Lovejoy DA, De Lannoy L. Evolution and phylogeny of the corticotropin-releasing factor (CRF) family of peptides: expansion and specialization in the vertebrates. *J Neurochem Anat*. (2013) 54:50–6. doi: 10.1016/j.jchemneu.2013.09.006
- Lovejoy DA, Jahan S. Phylogeny and Evolution of the corticotropin releasing factor family of peptides. *Gen Comp Endocrinol*. (2006) 146:1–8. doi: 10.1016/j.ygcen.2005.11.019
- Lovejoy DA, Chang B, Lovejoy N, Del Castillo J. Molecular evolution and origin of the corticotropin-releasing hormone receptors. *J Mol Endocrinol*. (2014) 52:6043–79. doi: 10.1530/JME-13-0238
- Coast GM. Insect diuretic peptides: structures, evolution and actions. *Amer Zool*. (1998) 38:442–9. doi: 10.1093/icb/38.3.442
- Kataoka H, Troetschler RG, Li JB, Kramer SJ, Carney RL, Schooley DA. Isolation and identification of a diuretic hormone from the tobacco hornworm, *Manduca sexta*. *Proc Natl Acad Sci USA*. (1989) 86:2976–80. doi: 10.1073/pnas.86.8.2976
- Zandawala M. Calcitonin-like diuretic hormones in insects. *Insect Biochem Mol Biol*. (2012) 42:816–25. doi: 10.1016/j.ibmb.2012.06.006
- Fredriksson R, Lagerström M, Lundin L, Schiöth H. The G-protein coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups and fingerprints. *Mol Pharmacol*. (2003) 63:1256–72. doi: 10.1124/mol.63.6.1256
- Nordström K, Lagerström M, Wallér L, Fredriksson R, Schiöth H. The secretin GPCRs descended from the family of adhesion GPCRs. *Mol Biol Evol*. (2009) 26:71–84. doi: 10.1093/molbev/msn228
- Schiöth H, Nordström M, Fredriksson R. The adhesion GPCRs: gene repertoire, phylogeny and evolution. In: Yona S, Stacey M, editors. *Adhesion GPCRs: Structure to Function*. Berlin: Landes Bioscience and Springer Science and Business Media (2010). p. 1–13. doi: 10.1007/978-1-4419-7913-1_1
- Lelianova V, Davletov B, Sterling A, Rahman M, Grishin E, Totty N, et al. A-latrotoxin, receptor, latrophilin, is a novel member of the secretin family of G protein-coupled receptors. *J Biol Chem*. (1997) 272:21504–8. doi: 10.1074/jbc.272.34.21504
- Holz GG, Habener JF. Black widow spider α -latrotoxin: a presynaptic neurotoxin that shares structural homology with the glucagon-like peptide-1 family of insulin secretagogic hormones. *Comp Biochem Physiol B Biochem Mol Biol*. (1989) 121:177–84. doi: 10.1016/S0305-0491(98)10088-3
- Silva JP, Lelianova VG, Ermolyuk YS, Hitchen PG, Berninhausen O, Rahman MA, et al. Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high affinity transsynaptic receptor pair with signaling capabilities. *Proc Natl Acad Sci USA*. (2011) 108:12113–8. doi: 10.1073/pnas.1019434108
- Boucard AA, Maxeiner S, Sudhof TC. Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: regulation by alternative splicing. *J Biol Chem*. (2014) 289:387–402. doi: 10.1074/jbc.M113.504779
- O'Sullivan ML, Martini F, von Daake S, Comoletti D, Ghosh A. LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5. *Neural Dev*. (2014) 9:7. doi: 10.1186/1749-8104-9-7
- Qian X, Baryshte-Lovejoy D, Wang L, Chewpoy B, Gautam N, Al Chawaf A, et al. Cloning and characterization of teneurin C-terminus associated peptide (TCAP)-3 from the hypothalamus of an adult rainbow trout (*Oncorhynchus mykiss*). *Gen Comp Endocrinol*. (2004) 137:205–16. doi: 10.1016/j.ygcen.2004.02.007
- Wang L, Rotzinger S, Al Chawaf A, Elias C, Baryshte-Lovejoy D, Qian X, et al. Teneurin proteins possess a carboxy terminal sequence with neuromodulatory activity. *Mol Brain Res*. (2005) 133:253–65. doi: 10.1016/j.molbrainres.2004.10.019
- Lovejoy DA, Al Chawaf A, Cadinouche A. Teneurin C-terminal associated peptides: An enigmatic family of neuropeptides with structural similarity to the corticotropin-releasing factor and calcitonin families of peptides. *Gen Comp Endocrinol*. (2006) 148:299–305. doi: 10.1016/j.ygcen.2006.01.012
- Jackson VA, Meijer DH, Carrasquero M, van Bezouwen LS, Lowe ED, Kleanthous C, et al. Structures of Teneurin adhesion receptors reveal an ancient fold for cell-cell interaction. *Nat Commun*. (2018) 9:1079. doi: 10.1038/s41467-018-03460-0
- Li J, Shalev-Benami M, Sando R, Jiang X, Kibrom A, Wang J, et al. Structural basis for teneurin function in circuit-wiring: a toxin motif at the synapse. *Cell*. (2018) 173:735–48. doi: 10.1016/j.cell.2018.03.036
- Husić M, Baryshte-Lovejoy D, Lovejoy DA. Teneurin C-terminal associated peptide (TCAP)-1 and latrophilin interaction in HEK293 Cells: evidence for modulation of intercellular adhesion. *Front Endocrinol*. (2019) 10:22. doi: 10.3389/fendo.2019.00022
- Tucker RP. Horizontal gene transfer in choanoflagellates. *J Exp Zool B Mol Dev Evol*. (2013) 320:1–9. doi: 10.1002/jez.b.22480
- Tucker R, Beckmann J, Leachman N, Schöler J, Chiquet-Ehrismann R. Phylogenetic analysis of the teneurins: conserved features and premetazoan ancestry. *Mol Biol Evol*. (2012) 29:1019–29. doi: 10.1093/molbev/msr271
- Zhang D, de Souza R, Anantharaman V, Iyer L, Aravind L. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct*. (2012) 7:18. doi: 10.1186/1745-6150-7-18
- Chand H, de Lannoy L, Tucker RP, Lovejoy DA. Origin of chordate peptides by horizontal protozoan gene transfer in early metazoans and protists: evolution of the teneurin C-terminal associated peptides. *Gen Comp Endocrinol*. (2013) 188:144–50. doi: 10.1016/j.ygcen.2013.02.006
- Endsin MJ, Michalec OM, Manzon LA, Lovejoy DA, Manzon RA. CRH peptide evolution occurred in three phases: evidence from characterizing sea lamprey CRH system members. *Gen Comp Endocrinol*. (2017) 240:162–73. doi: 10.1016/j.ygcen.2016.10.009
- Semmens DC, Mirabeau O, Moghul I, Pancholi MR, Wurm Y, Elphick MR. Transcriptomic identification of starfish neuropeptide precursors yields new insights into neuropeptide evolution. *Open Biol*. (2016) 6:150224. doi: 10.1098/rsob.150224
- Lovejoy DA, Baryshte-Lovejoy D. Characterization of a diuretic hormone-like peptide from tunicates: insight into the origins of the vertebrate corticotropin-releasing factor (CRF) family. *Gen Comp Endocrinol*. (2010) 165:330–6. doi: 10.1016/j.ygcen.2009.07.013

38. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA 6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* (2013) 30:2725–9. doi: 10.1093/molbev/mst197
39. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* (2004) 32:1792–7. doi: 10.1093/nar/gkh340
40. Weiss M, Steiner DF, Philipson LH. Insulin biosynthesis, secretion, structure and structure-activity relationships. In: De Groot LJ, Chrousos G, Dungan K, editor. *Endotext*. South Dartmouth, MA: NIH/NLM/NCBI Books (2014). p. 1–6.
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. Lang BF, O’Kelly C, Nerad T, Gray MW, Burger G. The closest unicellular relatives of animals. *Curr Biol.* (2002) 12:1773–8. doi: 10.1016/S0960-9822(02)01187-9
43. Minet A, Rubin B, Tucker R, Baumgartner S, Chiquet-Ehrismann R. Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene *Ten-m*, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci.* (1999) 112:2019–32.
44. Wimalawansa SJ. Amylin, calcitonin gene-related peptide, calcitonin and adrenomedullin: a peptide superfamily. *Critic Rev Neurobiol.* (1997) 11:167–239. doi: 10.1615/CritRevNeurobiol.v11.i2-3.40

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Michalec, Chang, Lovejoy and Lovejoy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.