Received: 18 June 2025 | Revised: 26 March 2026 | Accepted: 20 May 2026

DOI: 10.1111/jne.70203

ORIGINAL ARTICLE

Journal of Neuroendocrinology WILEY

The 14th World Congress on Neurohypophysial Hormones (WCNH)

Nonapeptide molecular evolution during the adaptive radiation of Tanganyika cichlids

Pol Sorigue[1,2] | Walter Salzburger[3] | Rui F. Oliveira[1,2]

[1]GIMM—Gulbenkian Institute for Molecular Medicine, Oeiras, Portugal

[2]ISPA—University Institute for Psychological, Social and Life Sciences, Lisbon, Portugal

[3]Department of Environmental Sciences, Zoological Institute, University of Basel, Basel, Switzerland

Correspondence
Rui F. Oliveira, GIMM—Gulbenkian Institute for Molecular Medicine, Oeiras, Portugal.
Email: rui.oliveira@gimm.pt

Funding information
Fundação para a Ciência e a Tecnologia, Grant/Award Number: PTDC/BIA-COM/3068/2020; Swiss national Science Foundation, Grant/Award Number: 208002

Abstract

Oxytocin (OT) and vasotocin (VT) are evolutionarily conserved nonapeptides that regulate a wide range of physiological and behavioral processes in vertebrates. Their receptor families have undergone gene duplications that facilitated functional diversification throughout vertebrate evolution. Using the diverse cichlid species in Lake Tanganyika, which have undergone repeated evolutionary transitions between social phenotypes, we investigated the molecular evolution of the nonapeptide system and its potential involvement in social behavior. We performed a positive selection analysis based on the *dN/dS* ratio and examined the correlation between amino acid variants and two social phenotypes. We also analysed gene expression data to explore associations between brain receptor expression and social phenotype variation. Our findings reveal that, while most sites in nonapeptide receptors are under strong purifying selection, a few sites– primarily in the extended intracellular loop 3 (IL3) of VTR2A receptors– show signatures of positive selection. Additionally, a specific amino acid in VTR2Aa correlates with pair-bonding, suggesting its potential role in social attachment. Gene expression analyses further revealed that components of the nonapeptide system, including VTR2Bb and OT, are differentially expressed across social phenotypes, supporting a role for regulatory variation alongside coding changes. Together, these findings provide new insights into how conserved neuroendocrine systems contribute to social diversity in cichlids.

KEYWORDS
cichlids, nonapeptides, oxytocin, social behavior, vasotocin

1 | INTRODUCTION

Oxytocin (OT) and vasotocin (VT) are two closely related nonapeptides that play essential roles in various physiological processes across vertebrates. Derived from a common ancestral molecule, these peptides share a highly conserved structure and function alongside two evolving families of receptors. Gene duplications have expanded receptor diversity, enabling the acquisition of new functions over evolutionary time. OT and VT have been extensively studied for their roles in osmoregulation, reproduction, and social behavior,[1,2] yet their precise mechanisms of action and evolutionary adaptations remain incompletely understood. In this study, we follow the universal nomenclature proposed by Theofanopoulou et al.[3] and supported by other studies,[4] which identifies VT as the ancestral vertebrate ortholog of mammalian arginine vasopressin (AVP), and OT refers to all oxytocin orthologs, including isotocin in teleosts. This nomenclature

reflects the evolutionary history of the genes and aids in cross-species comparisons.

An ancestral VT-like peptide and its cognate receptor existed before the emergence of vertebrates. This ancient VT-like peptide played key roles in the regulation of reproduction and in diuretic functions.[5,6] Around 540 million years ago (mya), the ancestral gnathostome vasotocin receptor (*VTR*) gene underwent a local duplication giving rise to a *VTR1/OT* receptor (*OTR*) and a *VTR2*. In turn, a segmental duplication of *VT* gave rise to *OT* around 530 mya,[7–9] leading to a contiguous head-to-tail genomic arrangement of the two ligand genes. This disposition is common across vertebrates, with the exception of mammals, in which the orientation is tail-to-tail so that the genes are transcribed from different strands; also, in some teleosts, the two nonapeptides are found on separate chromosomes.[8] The receptor evolution in vertebrates was determined by two rounds of whole genome duplication (1R, 2R) followed by clade-specific losses.[3,4] The teleost-specific third whole genome duplication (3R) event expanded the receptor repertoire in this group, which now includes two OTRs and up to seven VTRs, classified into four receptor types: VTR1A, VTR2A, VTR2B, and VTR2C.[3,4,10]

Nonapeptides are encoded by precursor genes of roughly 2000 base pairs that typically contain three protein-coding exons. These genes produce precursor proteins of more than 100 amino acids composed of a signal peptide, the hormone moiety, and a neurophysin domain, with an additional copeptin domain present in vertebrate VT and teleost OT.[11] The cyclic nonapeptide structure is highly conserved across vertebrates, with variation mainly at amino acid positions 3, 4, and 8. Based on the residue at position 8, nonapeptides are classified into basic (VT family) or neutral (OT family) peptides.[7] In teleosts, VT and OT sequences differ primarily at positions 4 and 8.[12] Neurophysin functions as a carrier protein with important post-translational roles,[13] while the copeptin domain has been suggested to play a role in prolactin release.[14] OTRs and VTRs are G protein coupled receptors (GPCRs) with seven transmembrane α-helices domains (TM1–7), alternated by three intracellular loops (IL1–3) and three extracellular loops (EL1–3). The N-terminal domain is exposed to the extracellular space and is involved in ligand binding, while the intracellular C-terminus interacts with the G protein for intracellular signal transduction.[15] Generally, OTRs and VTRs share high sequence similarity and several conserved functional residues across vertebrates.[12] While most nonapeptide receptors are encoded by two exons, the teleost *VTR2A* lineage evolved a distinct 5–4 exon–intron organization, giving rise to the paralogs *VTR2Aa* and *VTR2Ab*. Notably, VTR2A receptors possess an extended intracellular loop 3 (IL3) enriched in potential phosphorylation sites, suggesting increased regulatory and functional flexibility. Functionally, OTRs, VTR1s, VTR2Bs and VTR2C primarily couple with G_Q and activate the phospholipase C calcium-mobilization pathway,[16,17] whereas VTR2A couples with G_S and activates the cAMP signaling pathway.[18,19] Experimental evidence indicates that IL2 is required for G_Q activation, while IL3 is necessary and sufficient for G_S activation,[20] highlighting IL3 as a key component of VTR2A signaling.

Gene duplications in nonapeptide receptors facilitated functional diversification in vertebrates, enabling the evolution of novel ligand specificities, signaling pathways, and expression patterns.[21] Beyond their ancestral roles in reproduction and osmoregulation, the nonapeptide system (hereafter referring specifically to the OT/VT family, including OT, VT and their receptors) has been implicated in other roles,[1] including social behavior. In particular, this system has been implicated in the regulation of social behaviors. In mammals, studies in species such as the prairie vole (*Microtus ochrogaster*) have shown that variation in OT and vasopressin receptor distribution in reward-related brain regions is associated with differences in social bonding strategies.[22,23] Evidence from teleosts further supports a conserved role of the nonapeptide system in modulating diverse social behaviors.[24–26]

In teleosts, nonapeptide-secreting neurons are located in the preoptic area and project to multiple brain regions and the spinal cord.[27,28] The spatial distribution of receptors plays a central role in shaping nonapeptide activity and neural circuit function, and is therefore critical for behavioral modulation.[29,30] OTRs and VTR1As are broadly expressed within the teleost social decision-making network, an evolutionarily conserved system integrating social and reward-related processes.[31–34] Although VTR2s have also been detected in the brain,[35] their central functions are much less characterized. The nonapeptide system has been associated with a range of social behaviors in teleosts, including aggression, affiliation, parental care, and pair-bonding,[36–39] but species-specific patterns and receptor cross-binding contribute to functional complexity and limit the assignment of discrete roles.[40]

Cichlids are freshwater fishes known for their remarkable adaptive radiations, which have led to extensive speciation events.[41,42] The adaptive radiation of cichlid fishes in Lake Tanganyika comprises 243 endemic species, along with several riverine species.[43] These species are classified into 12 tribes and exhibit extensive phenotypic diversity, including variations in social behavior. Multiple independent evolutionary transitions between social phenotypes occurred, which makes this system ideal for comparative studies. A variety of mating systems have evolved within the Lake Tanganyika cichlid radiation, which can be broadly categorized as monogamous and polygamous, and these are tightly linked to caregiving strategies. Furthermore, the involvement of the nonapeptide system regulating social pair-bonding and parental care has been previously documented in cichlids.[38,44]

In this study, we investigate whether molecular evolution and gene regulation of the nonapeptide system are associated with the evolution of social phenotypes in the adaptive radiation of Lake Tanganyika cichlids. We focus on two repeatedly evolved social traits: (i) pair-bonding versus non-pair-bonding mating systems, and (ii) biparental versus maternal-only care. We first characterize the genomic repertoire and structural variation of nonapeptide ligands and receptors using whole-genome assemblies. We then test for signatures of positive selection across coding sequences and examine whether specific amino acid substitutions evolve in correlation with social phenotypes. Finally, we analyse tissue-specific gene expression and use phylogenetic generalized linear mixed models (PGLMMs) to assess whether variation in brain expression levels is associated with social behavior while accounting for shared evolutionary history.

Together, this integrative approach allows us to evaluate whether both coding and regulatory evolution of the nonapeptide system have contributed to the diversification of social systems in cichlids.

2 | METHODS

An extended version of the Methods section can be found in Supplementary Methods in Supplementary S2.

2.1 | Lake Tanganyika cichlids de novo assemblies

To reconstruct species-specific coding sequences of the nonapeptide system genes across the radiation, we generated de novo assemblies and species-level consensus sequences from available whole-genome data. Nile tilapia (*Oreochromis niloticus*, RefSeq accession GCF_001858045.2, female) was used as a reference genome because it is a well-annotated cichlid genome and phylogenetically equidistant to all Lake Tanganyika cichlid species. We used Illumina whole-genome raw reads from all cichlid species in Lake Tanganyika[41] to create de novo assemblies for all available species. Genomic raw reads (Bioproject PRJNA550295; NCBI BioProject database) were mapped to the reference genome using Bwa-mem (version 0.7.18[45];) with default parameters. In Bioproject PRJNA550295, genomes for all the species of the radiation are available, ranging from 1 to 4 individuals per species. Transcriptomic raw reads (Bioproject PRJNA552202[46];) were mapped to the reference genome using STAR aligner (version 2.7.10a[47];). All six tissues available (brain, gills, lower pharyngeal jaw, liver, ovaries, and testis) and 4 to 6 individuals across 74 species were used for the analysis. Detailed information on number of individuals per species and tissue availability is provided in Table S1.

2.2 | Nonapeptide system genes repertoire

To identify the full repertoire of nonapeptide precursors and receptor genes in the reference genome, we used zebrafish (*Danio rerio*, GenBank assembly GCA_000002035.4) and Japanese medaka (*Oryzias latipes*, GenBank assembly GCA_053564925.1), whose receptor repertoires are well characterized and span key teleost lineages, facilitating accurate receptor classification. Protein sequences of nonapeptides[7] and their receptors[4] were used to search the reference genome using protein BLAST ("BLASTp," https://blast.ncbi.nlm.nih.gov). Nile tilapia sequences matching the search (e-value <1e-50) were retained for further analysis (Table S2). Receptor types were assigned by building a phylogenetic tree with IQ-TREE (version 2.0.3[48];) using sequences from distantly related teleosts (Data S1).

Once the full repertoire of genes in the reference genome was identified, we investigated the presence and copy number of nonapeptide genes in Lake Tanganyika cichlids. To this end, we used all available long-read PacBio assemblies at the time of analysis from six species representing six tribes: *Neolamprologus multifasciatus* (Lamprologini; GenBank assembly GCA_963576455.2), *Simochromis diagramma* (Tropheini; GenBank assembly GCA_900408965.1), *Bathybates minor* (Bathybatini), *Cyphotilapia frontosa* (Cyphotilapiini), *Cunningtonia longiventralis* (Ectodini), and *Cyprichromis leptosoma* (Cyprichromini) (unpublished). Local BLAST databases were created with makeblastdb (version 2.16.0+), and BLASTn searches were performed using reference gene sequences. In all assemblies, a clear match spanning each gene was identified (e-value <1e-200; Table S3).

Because *dN/dS* analysis is based on protein-coding regions, we extracted coding sequences from Nile tilapia RefSeq annotations and included all splice variants with differences in protein sequence. Additional transcripts were identified through BLAST searches in NCBI ("African cichlids"), revealing one extra *OTRb* transcript. In total, 16 splice variants across 10 genes were retained. The coverage of all transcripts in all cichlid assemblies was calculated using Samtools depth.[49] Combining BLASTn results from long-read assemblies with Illumina coverage data, we identified a 16 bp deletion in exon 4 of VTR2Ab in multiple Lamprologini species, along with four additional features (Supplementary Methods in Supplementary S2). The consensus sequence of each gene was obtained for each species using the genomic location of the coding sequences and the BAM files. We obtained a species consensus by merging the individual genes of the same species with Seqtk (https://github.com/lh3/seqtk/). To obtain the consensus from the BAM file, we used a combination of SAMtools and BCFtools.[49] Protein domains were annotated using InterPro (https://www.ebi.ac.uk/interpro/search/sequence), providing amino acid positions and domain architecture across sequences.

2.3 | Multiple sequence alignments

We built a multiple sequence alignment with MAFFT (version v7.526[50];) for each gene with the consensus sequences of all Lake Tanganyika cichlids and the reference gene from Nile tilapia genome. We used genomic and transcriptomic assemblies to improve the accuracy of the consensus. We then translated every coding sequence into amino acids and checked for the presence of START/STOP codons, early STOP codons or sequence lengths not divisible by three. We spotted multiple STOP codons with the sequence "TGA." Although the codon "TGA" is known as a translation STOP sequence, this also encodes a selenocysteine and does not necessarily stop the translation.[51] *VT* and *VTR2Bb.tr1* did not have a canonical START and we intentionally trimmed the alignment until the first methionine. Few additional manual changes were applied (discussed in Supplementary Methods in Supplementary S2). After the manual modifications, a codon-based alignment was then performed using PAL2NAL (version V14[52];) ensuring that positive selection analysis was computed on the correct reading frame.

We calculated the nucleotide diversity index (π) of the alignments using the Ape package (version 5.8) in R (version 4.4.0). To examine domain-specific variation, we calculated π for all domains only in transcripts that encode a canonical GPCR. We first calculated π for each of the individual domains and then we computed π grouping domains

by type. Additionally, in order to spot possible evolutionary constraints in the genes, we calculated the rate of synonymous substitutions for all transcripts and all species using Yang and Nielsen method.[53]

2.4 | Positive selection

To test whether protein-coding evolution of nonapeptide genes shows signatures consistent with adaptive divergence, we performed gene-wide and site-specific *dN/dS* analyses. For each transcript, we built a ML phylogenetic tree using IQTREE (version 2.0.3[48];). Among different substitution models, IQTREE automatically selects the best fit and creates a phylogenetic tree. We built gene trees with nucleotide and amino acid sequences. Whole-gene ω (*dN/dS*) values were obtained using HyPhy (version 2.5.62) with the Fixed Effects Likelihood (FEL) method.[54] We also used FEL to investigate the site-specific positive selection across the gene sequences of the nonapeptide system genes. Because the multiple sequence alignments showed variable dS values (Data S3), we selected FEL as an appropriate method due to its ability to account for dS variation across sites and branches.[55] We accepted a *p*-value of <.05 as a threshold of significance. We performed positive selection using gene trees and species trees for all transcripts.

2.5 | Correlated evolution between discrete traits

To determine whether social phenotypes evolved in a correlated manner across the radiation, we used comparative discrete trait evolution models. We gathered phenotypic data for all species for which information was available in our lab. We gathered data on the mating system for 175 species: 105 that form pairs and 70 that do not. We obtained phenotypic information on the caregiving parent for 212 species: 86 species with biparental care and 126 that carry out maternal care.

To assess correlated evolution between traits, we used the "discrete method" within the program BayesTraits (version V4.0[56];). Each trait was binarized for compatibility reasons. We run two different models: an independent model, where the two traits evolved separately, and a dependent model, where the evolution is correlated. We applied a Markov-Chain Monte Carlo (MCMC) with uniform priors to estimate Bayesian posterior distributions for rate parameters, and the "stepping stone" sampler to obtain marginal log-likelihoods. To compare the models, we calculated the Log Bayes Factor (Log BF) as described in the manual. We used the reversible-jump approach to assess precedence and contingency between the two traits.[57] We sampled a gamma prior with both parameters (shape and scale) ranging between 0 and 100, seeded from a uniform hyperprior ranging from 0 to 100. To test convergence, we applied Geweke's convergence diagnostic in the Coda package (version 0.19-4.1) in R. For a detailed description of this section, see Supplementary Methods in Supplementary S2.

2.6 | Tissue expression and correlation with behavioral phenotypes

To study the tissue expression distribution of the nonapeptide precursors and receptors in Lake Tanganyika's cichlids, we used the CPM-normalized count dataset from El Taher et al.[46] Bulk RNA of six tissues (brain, liver, lower pharyngeal jaw, gills, ovaries, and testis) of males and females of 74 species is available.

To test for associations between brain gene expression and behavioral phenotypes while accounting for shared evolutionary history, we fitted PGLMMs for each gene independently using the R package MCMCglmm (version 2.36). Brain gene expression values were counts-per-million normalized, $\log_2$-transformed and scaled to stabilize variance and improve comparability across genes. Fixed effects included the behavioral phenotype of interest, sex, their interaction, and stable isotope values linked to diet ($\delta^{15}N$) and habitat ($\delta^{13}C$). To account for phylogenetic non-independence among species, species identity was included as a random effect with a covariance structure derived from the Lake Tanganyika phylogeny. The inverse phylogenetic relatedness matrix was supplied to the model using the ginverse argument. Only species with available phenotype data were included in each analysis. Models were fitted assuming Gaussian error distributions and were run for 1,000,000 MCMC iterations, with a burn-in of 10,000 iterations and a thinning interval of 100, yielding 9900 posterior samples per parameter. Convergence was assessed using Geweke diagnostics for all fixed effects, and a threshold of $|z| < 1.95$ was considered indicative of adequate convergence. Effective sample sizes (ESS) were calculated for all fixed effects. Posterior distributions were used to estimate effect sizes (posterior means), two-tailed Bayesian *p*-values (pMCMC; calculated as twice the minimum posterior tail probability), and 95% highest posterior density (HPD) intervals for each fixed effect.

3 | RESULTS

3.1 | Nonapeptide system genes repertoire

Both nonapeptide precursor genes—OT and VT—and eight receptors were found in the Nile tilapia genome (*Oreochromis niloticus*; RefSeq accession GCF_001858045.2, female). To assign a receptor type to each sequence, we built a phylogenetic tree with known gene sequences from distant teleost species. Based on sequence similarity with multiple teleost species, we assigned the following repertoire of genes in Nile tilapia: two *OTR*s, two *VTR1A*s, two *VTR2A*s and two *VTR2B*s (Table S2). The species tree and sequences used can be consulted in Data S1.

To find the gene repertoire in Lake Tanganyika cichlids, we performed a nucleotide BLAST search with the genes of the reference genome on all six long-read assemblies. Despite some genetic differences between Lake Tanganyika cichlids and the reference genome, we found a single clear match for each reference gene (e-value <1e-200). Using the location on the Nile tilapia chromosomes, we

compared the distribution of the genes across the long scaffolds in the six long-read assemblies (Table S3). Both nonapeptide precursor genes were found on the same scaffold in a head-to-tail conformation and separated by 40–50 kbp, consistent with the annotation in Nile tilapia (Figure 1). The predicted nonapeptide amino acid sequences in Lake Tanganyika cichlids follow the canonical teleost oxytocin (OT: CYISNCPIG; Cys-Tyr-Ile-Ser-Asn-Cys-Pro-Ile-Gly) and vasotocin (VT: CYIQNCPRG; Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Arg-Gly) sequences. The location of all nonapeptide receptors in the long-read assemblies was also consistent with the chromosomal distribution in the Nile tilapia genome (Figure 1). For positive selection analysis, we selected splicing forms that differed to any extent in their protein-coding sequences. A total of 15 splicing forms across the 10 genes were taken from the reference genome, and one additional transcript was included from other cichlid annotations (Figure 1). Except for a 66 bp deletion in the *VTR2Ab* gene, all genes showed a highly conserved protein-coding domain structure.

3.2 | Structure of the nonapeptides and their receptors

The predicted nonapeptide precursors preproteins followed the known conserved structure that includes a signal peptide, hormone moiety, GKR residue, neurophysin, and copeptin. All receptor genes had at least one isoform that exhibited the classic seven transmembrane (TM) domain structure, with three ILs and three ELs alternating between the TM domains (Figure 2A). Some alternative splicing forms (*OTRb.tr1*, *OTRb.tr3 and VTR2Ab.tr3*) deviate from this canonical GPCR structure, resulting in shorter isoforms with fewer TM domains and therefore fewer loops (Figure 2B–D). Additionally, we confirmed that the 66 bp deletion of the *VTR2Ab* gene was located in the extended IL3. A detailed predicted domain structure of each transcript can be consulted in Data S2.

3.3 | Nucleotide diversity of the nonapeptide system genes

OTRa exhibited the lowest nucleotide diversity (π) among all genes, followed by *OTRb*, *VTR1Aa*, and *VTR2Ba*. Both nonapeptide precursors displayed intermediate levels of variability, while *VTR2A* copies showed the highest diversity values (Data S3).

Across domains, we only analysed the isoforms with a classic GPCR structure. Terminal regions appeared to be the most variable, while TM domains and both extracellular and intracellular loops were less variable (Figure 3). Interestingly, the two extended IL3s (from *VTR2Aa* and *VTR2Ab* genes) showed the largest nucleotide diversity across all IL3. In addition, the C-terminal domains of both *VTR2A* copies are also among the most variable (Figure 3). The alignments and nucleotide diversity values can be consulted in Data S3.

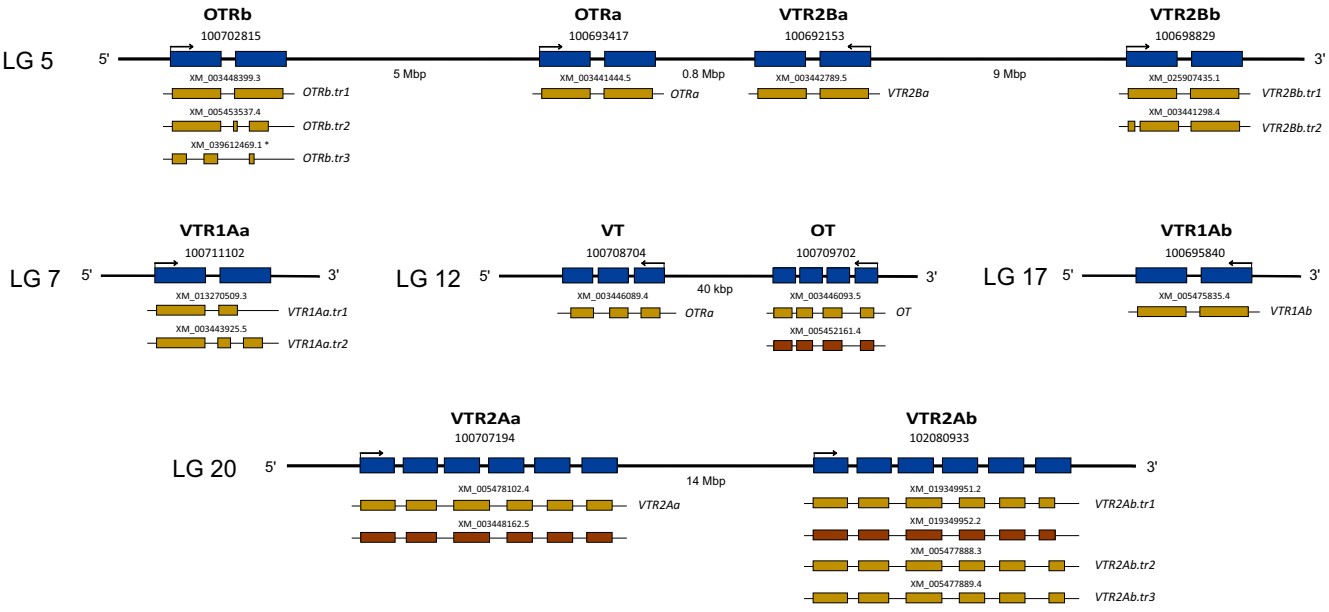

FIGURE 1 Genomic distribution of the nonapeptide system in the reference genome *Oreochromis niloticus*. In the present study we found evidence that suggests a similar distribution of the nonapeptide system genes in Lake Tanganyika cichlids. Gene IDs and transcript accession numbers are shown. All accession numbers were obtained from NCBI from Nile tilapia's genome assembly GCF_001858045.2, except for *XM_039612469.1*. At the right of each transcript, the name of the transcript is given as it is referred to in this study. In blue, schematic representation of the exons comprising the 10 genes. In gold, transcripts selected for the present study. In brown, transcripts that did no differ from another transcript in their protein-coding sequence and therefore were not taken for analysis. The arrows indicate the sense of transcription. The genes are sorted by chromosome (linkage group, LG) number and location. (*) This transcript is not annotated in the reference genome, the accession number belongs to an *OTRb* transcript of another cichlid, the blue tilapia (*Oreochromis aureus*).

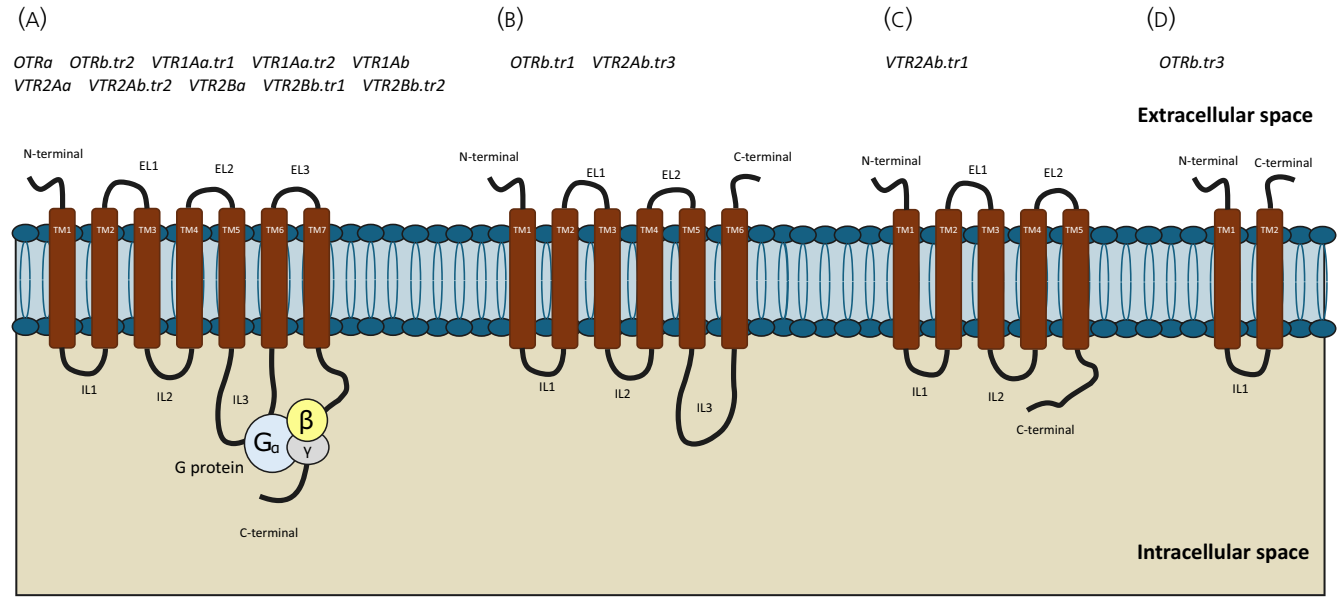

FIGURE 2 Schematic representation of the predicted domain structures of different alternative splicing variants of nonapeptide receptors in Lake Tanganyika cichlids. (A) Canonical G-protein-coupled receptor (GPCR) structure with seven transmembrane (TM) domains, three extracellular loops (EL1–EL3), and three intracellular loops (IL1–IL3), interacting with the G-protein complex. (B) *OTRb.tr1* and *VTR2Ab.tr3* isoforms, which lack the seventh TM domain, resulting in an extracellular C-terminus. (C) *VTR2Ab.tr1* isoform, consisting of five TM domains, two ELs, and two ILs. (D) *OTRb.tr3* isoform, which spans only two TM domains and has a single IL.

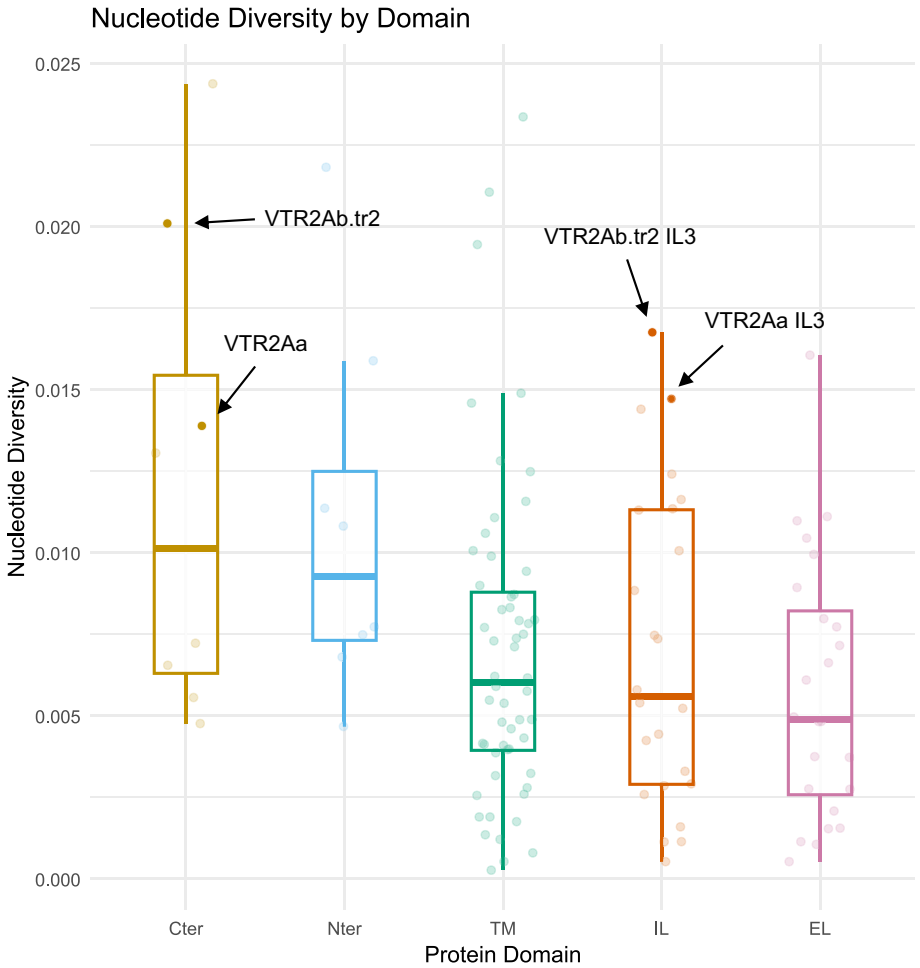

FIGURE 3 Nucleotide diversity (π) across protein domains of nonapeptide receptors. For each receptor gene, we selected one isoform with a predicted canonical G-protein-coupled receptor (GPCR) structure. Each dot represents the nucleotide diversity (π) estimate for a single gene within the indicated protein domain. Boxplots show the distribution of π values across genes: the central horizontal line indicates the median; box limits represent the interquartile range (IQR, 25th–75th percentiles); whiskers extend to 1.5 × IQR. Cter, C-terminus; Nter, N-terminus; TM, Transmembrane domain; IL, Intracellular loop; EL, Extracellular loop.

3.4 | Positive selection on nonapeptide system genes

Phylogenetic trees were inferred using IQTREE with automatic substitution model selection. The best-fitting nucleotide and amino acid substitution models varied among transcripts and are summarized in Table S4.

The ratio of nonsynonymous to synonymous substitution rates (*dN/dS*, ω) was used to assess patterns of molecular evolution in the analysed genes. This ratio compares the rate of amino acid–changing substitutions (*dN*) to that of synonymous substitutions that do not alter the encoded amino acid (*dS*). Conventionally, ω <1 indicates purifying selection, $\omega = 1$ is consistent with neutral evolution, and ω >1 suggests positive selection. However, these interpretations should be treated with caution when calculated across entire coding sequences, as positive selection typically acts on a limited number of codons while most of the gene remains under strong purifying selection. As a result, whole-gene ω values often remain below 1 even when positive selection is present at specific sites. Accordingly, we use ω primarily for comparative purposes rather than relying on its absolute values.

Varying values of synonymous substitutions were found across genes and transcripts (Table S1). Gene-wide ω (*dN/dS*) values show generalized purifying selection across all genes. *OTRa* had the lowest ω ($\omega = 0.1$ in both tree topologies), while both nonapeptides displayed relatively high values (Table 1). Both tree topologies yielded similar results.

Similarly, both tree topologies provided largely similar results in site-specific positive selection analysis. The following description is based on the gene tree topology, as we considered that it better represents the evolutionary history of the gene. All genes and transcripts were subject to strong negative selection (Figure 4; Data S4). A total of 335 sites under negative selection and 15 sites under positive selection (Table S4). Across the receptors, almost half the negatively selected sites (42%) belonged to TM domains, followed by ILs (26%) and ELs (16%).

No positively selected site was detected on the nonapeptide precursor genes. Similarly, *OTRa* showed no positively selected sites, and *OTRb* only one (p-value = .046), unique to a specific isoform, located in the C-terminal extracellular region. Two positively selected sites were detected in *VTR1Aa*; one in EL3 (p-value = .022), and one in the intracellular C-terminal region (p-value = .045), while *VTR1Ab* showed no positive selection. In contrast, *VTR2Aa* had two positively selected sites, with one site in IL3 being highly significant (p-value .0004). *VTR2Ab* exhibited the highest number of positively selected sites, with six in total, including four in the extended IL of its canonical copy *VTR2Ab.tr2* (Figure 5). In *VTR2Ba*, no positively selected sites were found under the gene tree topology. A detailed list of the sites under selection can be consulted in Data S4.

TABLE 1 Gene-wide ω values obtained for each transcript using the Fixed Effects Likelihood in HyPhy (version 2.5.62). The values are shown for the analysis performed with gene trees and species trees. Average genome-wide ω values are around 0.2.[41]

	Gene tree	Species tree
OT	0.384725	0.446579
VT	0.389876	0.378217
OTRa	0.103896	0.098344
OTRb.tr1	0.364471	0.396942
OTRb.tr2	0.320607	0.319785
OTRb.tr3	0.393161	0.405904
VTR1Aa.tr1	0.265012	0.250758
VTR1Aa.tr2	0.272318	0.256706
VTR1Ab	0.135947	0.138472
VTR2Aa	0.172017	0.17183
VTR2Ab.tr1	0.356552	0.346824
VTR2Ab.tr2	0.227011	0.223489
VTR2Ab.tr3	0.246962	0.277689
VTR2Ba	0.2334	0.198095
VTR2Bb.tr1	0.398828	0.331528
VTR2Bb.tr2	0.317791	0.260721

3.5 | Correlation between brooding and pair-bonding phenotypes

We found the two traits strongly correlated (log BF 84.36) in Lake Tanganyika adaptive radiation (Figure 6). According to our reversible-jump analysis, the ancestral state in Lake Tanganyika cichlids was a biparental species with a very high probability (~92%), and very likely a pair-bonding species (~74%). Together, the probability of a biparental and pair-bonding species was 70.42%. The evolution of these binary traits seemed to follow a direction toward maternal and non-pair-bonding species. Pair-bonding behavior seems to be easily lost when caregiving is only maternal. Instead, gaining pair-bonding behavior seems to be very unlikely (Data S5).

3.6 | Correlation between behavioral phenotypes and SNPs

Of the 1098 sites tested for correlation with both phenotypes, only one amino acid showed evidence of evolution associated with pair-bonding: position 369 of the VTR2Aa transcript (green amino acid in Figure 5). Under uninformative priors, this site yielded a log BF of 2.5 (Data S5) and stands out among the other sites (Figure S2). The SNP corresponds to a valine-to-alanine substitution. Given the large number of sites screened, this initial signal could reflect stochastic variation. We therefore reanalysed this candidate site using the reversible-jump framework implemented in BayesTraits with priors selected following the developer's recommendations. Under this approach, the dependent model was favored over the independent model with a log BF of 6.67, indicating strong support for correlated evolution according to the BayesTraits guidelines. The analysis further suggested a constraint

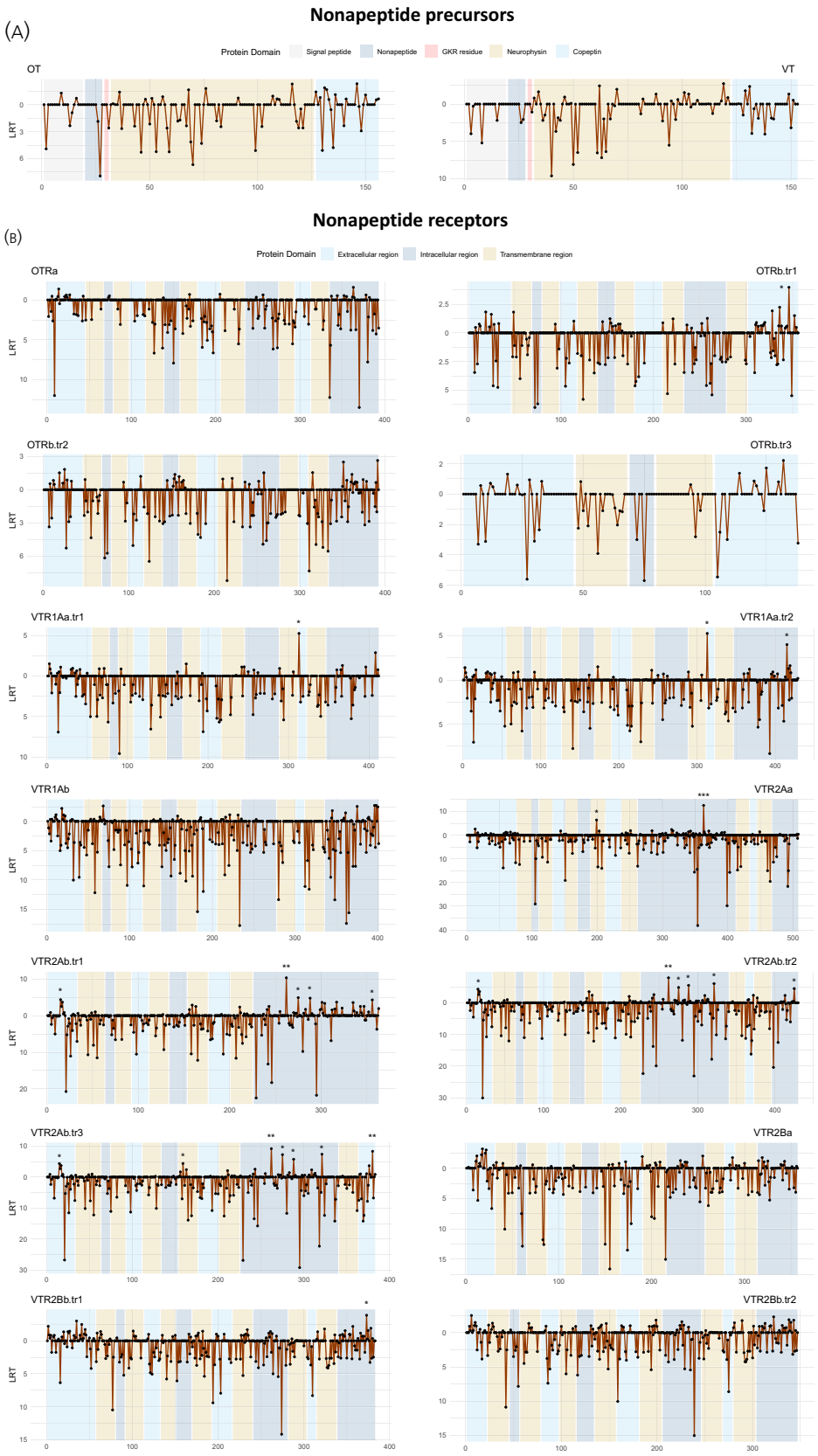

FIGURE 4 Likelihood ratio test (LRT) values of positive selection analysis along the protein sequences of nonapeptide precursors and receptors. *X*-axis represents amino acid number (site) along the protein sequence. LRT values represent the strength of selection at each codon position. For visualization purposes, LRT values are positive in both directions: Upward deviations indicate positive selection, and downward deviations indicate negative selection. Each plot corresponds to a different transcript. Background shading represents protein domains. Asterisks indicate significance in positively selected sites (*) *p*-value <.05 (**) *p*-value <.005 (***) *p*-value <.0005.

involving alanine in pair-bonding species. Valine, the predominant amino acid at this position, occurs in both pair-bonding and non–pair-bonding species, whereas alanine is predominantly associated with pair-bonding taxa. The likelihood values, as well as convergence diagnostic values and reversible-jump files can be consulted in Data S5.

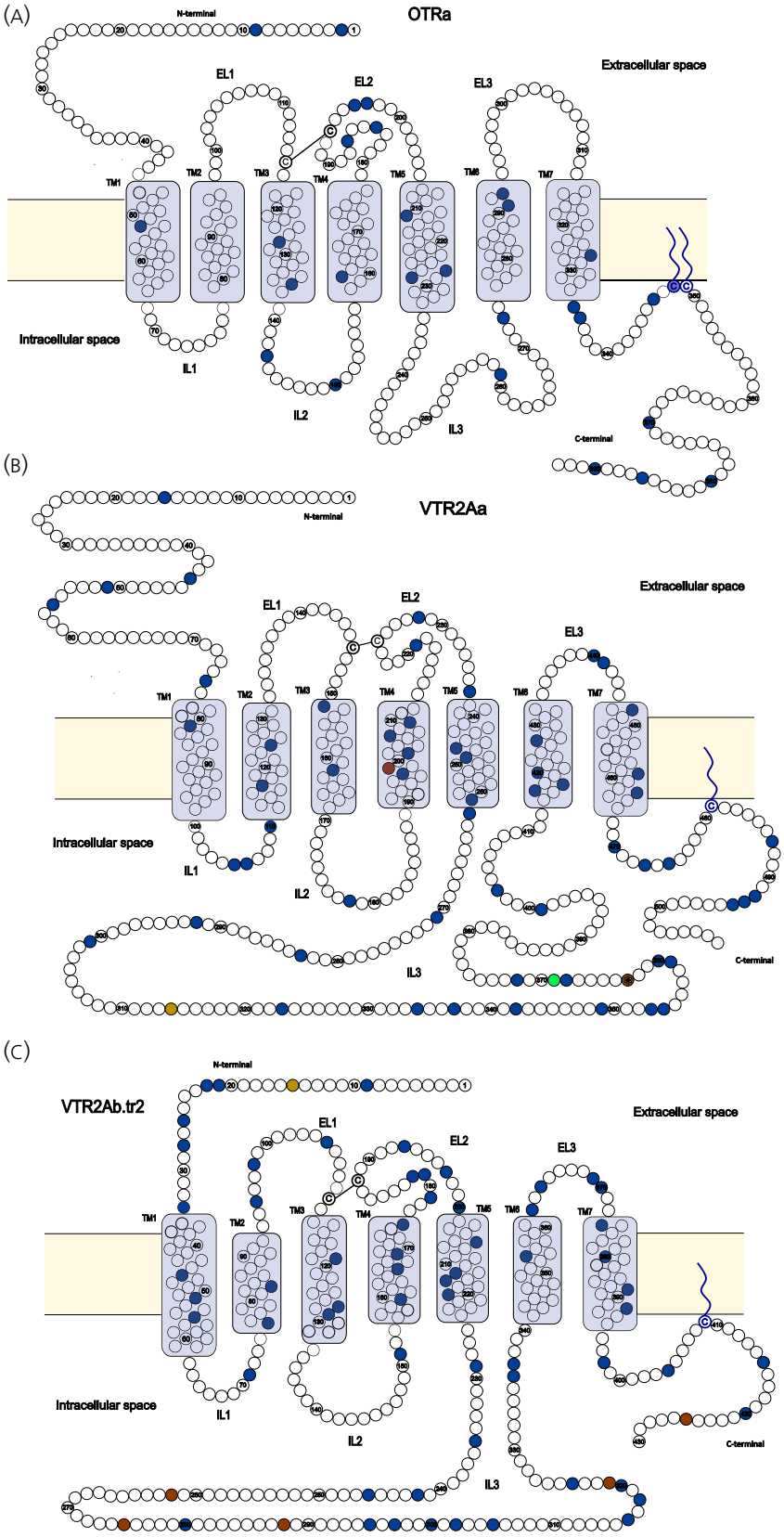

FIGURE 5 Schematic representation of nonapeptide receptor protein sequences showing selection analysis derived from *dN*/*dS* ratio: (A) *OTRb*, (B) *VTR2Aa*, (C) *VTR2Ab.tr2*. Dark blue circles indicate significant negatively selected sites found with any of the two tree topologies (gene tree and species tree). Golden circles indicate significant positive selection found only with species tree (*p*-value <.05). Brown circles indicate significant positive selection found with both tree topologies (*p*-value <.05). Asterisk indicates very significant positive selection (*p*-value <.0005 when calculated with the gene tree topology). The green circle indicates an amino acid that showed correlated evolution with pair-bonding behavior.

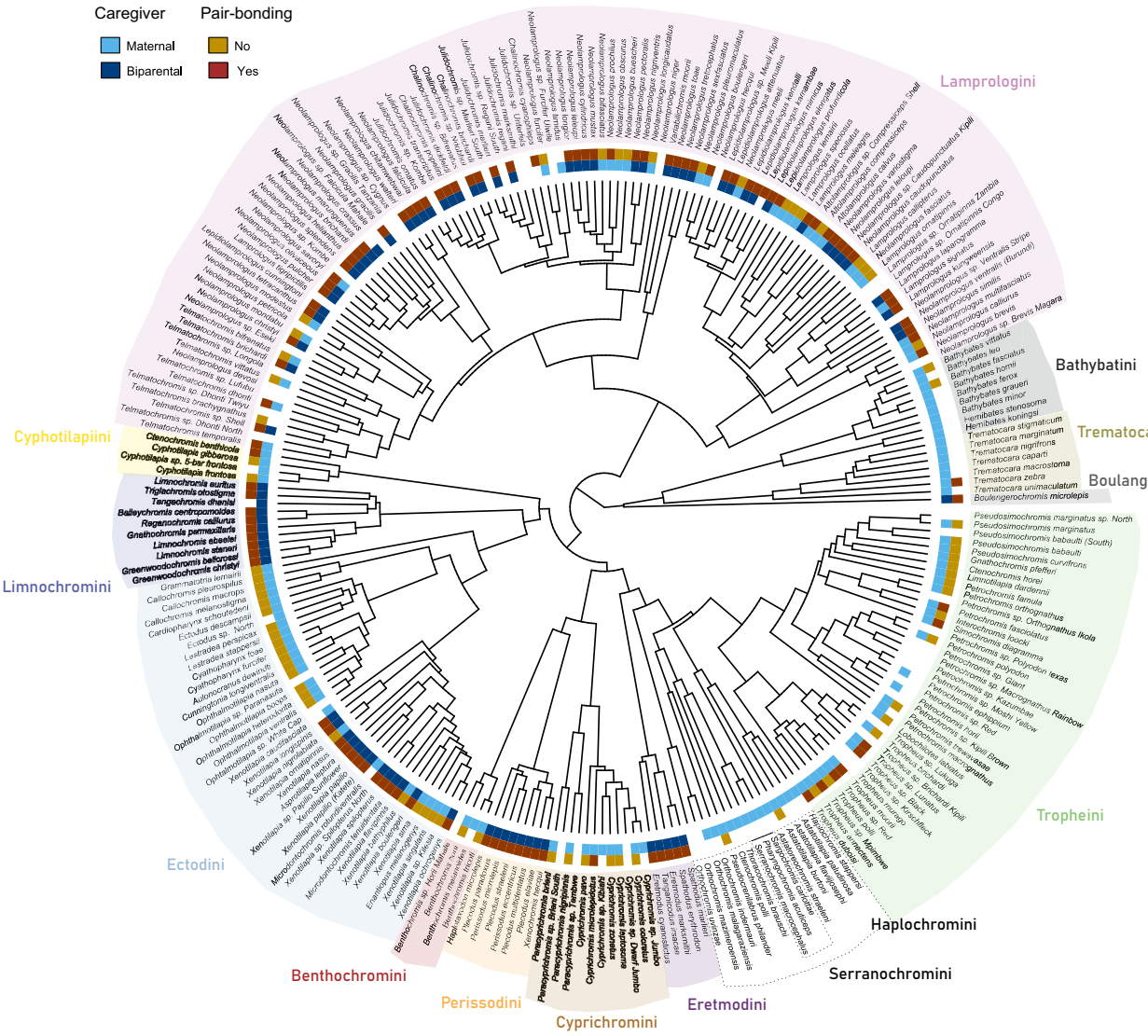

FIGURE 6 Phenotypic characterization of the Lake Tanganyika cichlid radiation based on caregiver sex and pair-bonding behavior. The circular phylogeny represents the relationships among species included in this study. Colored tiles surrounding the tree indicate discrete phenotypic states for each species. The inner ring corresponds to caregiver type (light blue = maternal care; dark blue = biparental care), and the outer ring corresponds to pair-bonding behavior (yellow = no pair-bonding; red = pair-bonding present). Each tile represents the state for a single species at the tip of the phylogeny. Background shading delineates major tribes within the Lake Tanganyika radiation (tribe names indicated on the outer margin). Multiple independent transitions between caregiver types and pair-bonding states are apparent across the phylogeny.

3.7 | Tissue expression and correlation with behavioral phenotypes

The nonapeptide precursors were almost exclusively expressed in the brain, their primary site of production. While *OTRa* was mainly expressed in the brain, *OTRb* showed very high expression in the gills. *VTR1Aa* was expressed in gills, lower pharyngeal jaw and testis, and to a lesser extent in the brain. The expression levels of *VTR1Ab* were generally lower and reduced to the gills and brain. All *VTR2s* were expressed in the brain (Figure 7). While the function of *VTR2s* in the brain remains less studied, their central expression appears to be consistent across teleosts.[35,58] *VTR2Aa* expression was widespread across different tissues, suggesting a role in multiple physiological functions. Its high expression in the liver aligns with previous reports linking it to carbohydrate metabolism.[59] Interestingly, its expression in the lower pharyngeal jaw was very high and its physiological role remains an open question. *VTR2Bb* showed expression in the ovaries and testis, in line with other studies.[60]

3.8 | Association between the expression of nonapeptide genes and behavioral phenotypes

All models showed adequate MCMC performance based on Geweke diagnostics (|z| <1.95) and high ESS across fixed effects, indicating reliable posterior estimation. A single exception was detected for the

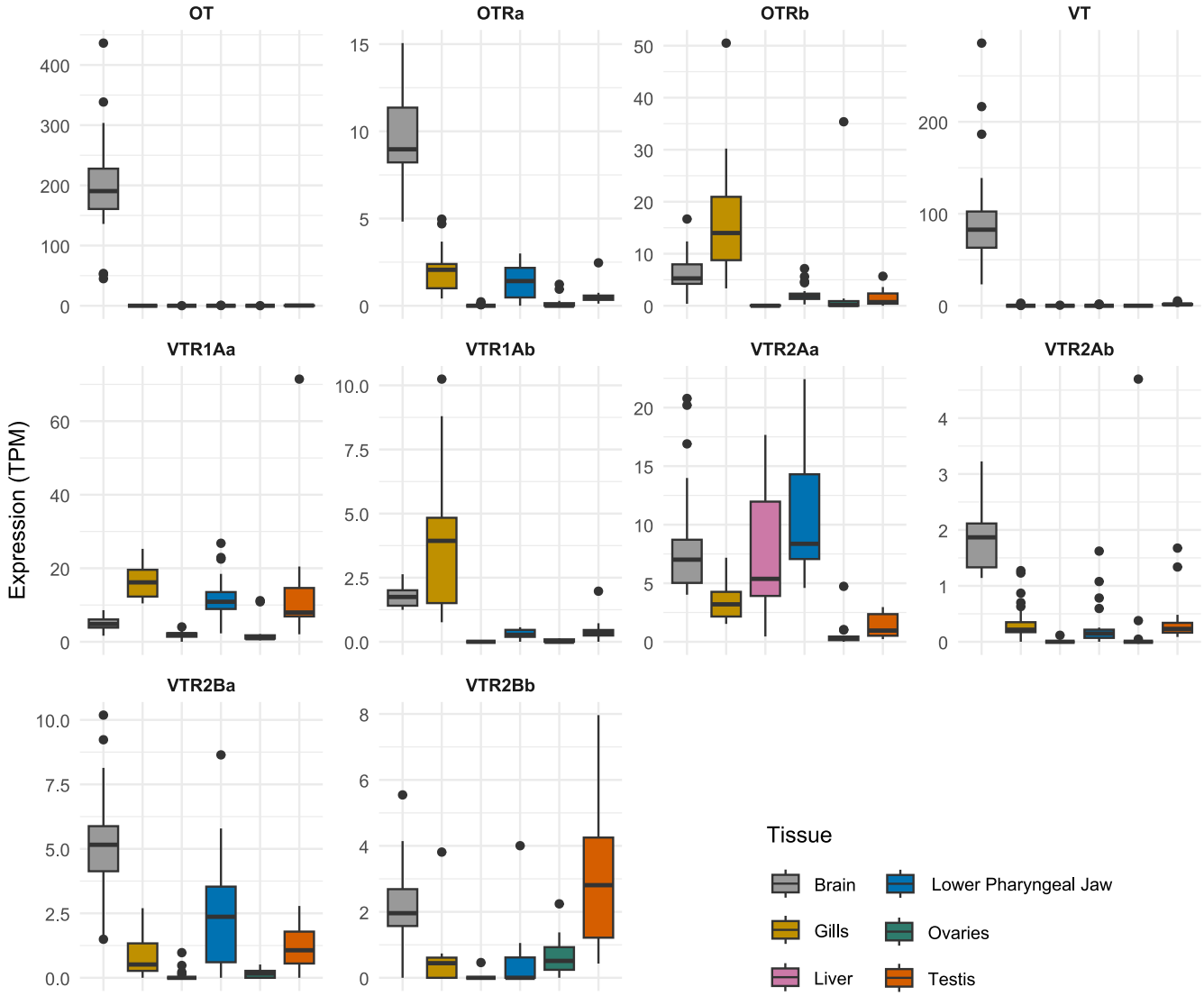

FIGURE 7 Tissue-specific expression patterns of nonapeptide system genes across six tissues. Each panel represents one gene of the nonapeptide system. Gene expression levels are shown as transcripts per million (TPM) obtained from whole-tissue bulk RNA sequencing. Boxplots summarize expression distributions across species for each tissue: the central horizontal line indicates the median; box limits represent the interquartile range (IQR, 25th–75th percentiles); whiskers extend to 1.5 × IQR; and points beyond whiskers represent outliers. Each colored box corresponds to a tissue type (Brain, Lower Pharyngeal Jaw, Gills, Ovaries, Liver, Testis; see legend). For visualization purposes, males of *Boulengerochromis microlepis* were excluded from oxytocin (OT) and vasotocin (VT) panels due to exceptionally high precursor expression values that strongly skewed the scale; inclusion of these values does not alter the overall qualitative tissue-specific patterns (see Section 2).

$\delta^{13}C$ effect in the VTR2Aa model when testing pair-bonding, where the Geweke diagnostic slightly exceeded the convergence threshold, suggesting incomplete mixing for that parameter ($|z| = -2.57$). All other parameters across genes and phenotypes showed satisfactory convergence. Minimum ESS across all fixed effects, genes, and datasets was 8984.49, indicating sufficient independent posterior samples across all parameters. We considered associations statistically supported when pMCMC <0.05 and the 95% HPD interval did not overlap zero.

VTR2Bb expression was substantially higher both in pair-bonding species (posterior mean = 0.5, 95% HPD = [0.13, 0.9], pMCMC = 0.011) and in biparental caregivers (posterior mean = 0.45, 95% HPD = [0.05, 0.84], pMCMC = 0.027). *OT* expression was higher in biparental species (posterior mean = 0.45, 95% HPD = [0.11, 0.82], pMCMC = 0.01). Sex-specific expression patterns were consistently observed in *OT*, *VT*, and *VTR2Aa*. While both ligands showed much higher expression in males (*OT*: posterior means = 0.62 and 0.57, pMCMC = 0 in both phenotype models; *VT*: posterior means = 0.39 and 0.4, pMCMC = 0.01 and 0.001), *VTR2Aa* showed increased expression in females (posterior means = −0.34 and −0.18, pMCMC = 0 and 0.015). Stable isotope effects were generally weak, with little evidence for associations across genes. The only notable pattern was lower *VTR2Bb* expression with increasing $\delta^{13}C$ (posterior means = −0.20 and −0.23, pMCMC = 0.03 and

0.01), and higher expression of *VTR1Aa* with increasing $\delta^{13}C$ (posterior means = 0.19 and 0.16, pMCMC = 0.03 and 0.04). Detailed model outputs, including full posterior summaries and diagnostics, are provided in Table S5. The CPM-normalized count dataset used for these analyses is available in Data S6.

4 | DISCUSSION

We characterized the nonapeptide system gene repertoire in the adaptive radiation of cichlid fishes from Lake Tanganyika. We found that these cichlids possess one copy of each nonapeptide precursor, separated by 40–50 kbp, with conserved teleost amino acid sequences for OT (CYISNCPIG) and VT (CYIQNCPRG). Long-read assemblies indicate that receptor gene copy number, genomic arrangement, and exon structure are largely conserved and comparable to those of Nile tilapia. *VTR2Ab* experienced a 66 bp deletion at an early stage in the radiation—after the divergence of Boulengerochromini, Bathybatini and Trematocarini lineages—that trims 22 amino acids of the IL3. A new early splicing site was acquired and the current widespread expression of this new variant evidences its functionality.

OTRb and *VTR2Ab* have alternative splicing forms whose predicted protein domains deviate from the canonical GPCR structure (7 TM domains alternating between 3 EL and 3 IL). The physiological function of the alternatively spliced receptor isoforms remains an open question. Notably, OTRb.tr1 and VTR2Ab.tr3 lack the final TM domain, resulting in an extracellular C-terminus. This structural alteration has been reported to prevent receptor migration to the plasma membrane, retaining it within the ER-Golgi compartments.[61] Although all three EL contribute to ligand binding, EL1 and EL2 appear to be essential for nonapeptide-receptor interactions: EL1 binds the tripeptidic part of the nonapeptide, while EL2 interacts with the cyclic part.[12,62,63] Therefore, OTRb.tr1, VTR2Ab.tr3, and VTR2Ab.tr1 may still retain their ligand-binding capacities, although potentially with altered affinity. In contrast, OTRb.tr3, which consists of only two TM domains and a single IL, possesses extracellular domains limited to the N- and C-terminal regions. This structural configuration makes direct interaction with the ligand far less likely, raising questions about its functional role. These findings suggest that alternative splicing may introduce additional layers of regulatory or functional diversity within an otherwise conserved system.

With the emergence of vertebrates (~540 mya) and two associated whole genome duplication events, new copies of VTRs acquired new functions. The neofunctionalization of these receptors is very well established in mammals,[2] but growing evidence is revealing the implication of nonapeptides in a wide range of functions in fish,[1] including social phenotypes such as aggression, pair-bonding, or parental care in teleost.[36–39] For neofunctionalization to occur, genetic diversification is necessary to introduce functional changes. To explore this possibility, we tested for positive selection in the nonapeptide system within the Lake Tanganyika cichlid adaptive radiation. We found that most amino acids were under strong negative selection, reflecting the strong evolutionary pressure to preserve protein structure and functionality. These negatively selected amino acids are distributed across all functional domains, consistent with the overall high conservation of the receptors across vertebrates.[4,10,12] Moreover, because of the young age of the radiation there are several sites that do not show any variation, which impedes making inferences about selection. 15 sites showed signs of positive selection when analysed with the gene tree topology. Interestingly, five of these sites were concentrated in the extended IL3 of *VTR2Aa* and *VTR2Ab*. In our analysis, the two extended IL3 showed the largest nucleotide diversity among all IL. Furthermore, multiple or highly significant amino acids in this domain showed positive selection. Taken together, our findings suggest that the overall structure of *VTR2Aa* and *VTR2Ab* is under purifying selection, similarly to the other nonapeptide receptors, while the part in charge of signal transduction is subject to change. At some point in teleost evolution, the intracellular loop (IL3) of VTR2As expanded significantly while presumably retaining receptor functionality.[4] This enlargement introduced additional amino acids into the domain responsible for signal transduction, directly interacting with the G protein. While some residues remain highly conserved and under strong negative selection, reduced constraint in this extended region may allow for greater evolutionary flexibility, potentially facilitating functional diversification or adaptation to different regulatory mechanisms. This extended IL3 has drawn very little attention to date. However, IL3 is necessary and sufficient for activation of G_s,[20] making it a crucial component of VTR2A downstream signaling. Further work should elucidate the functional implications and determine whether the extended IL3 contributes to specific regulatory mechanisms, signaling efficiency, or interactions with other intracellular partners.

In the cichlid adaptive radiation of Lake Tanganyika, in a relatively short period (~10 mya) multiple social phenotypes have repeatedly evolved, making it an ideal scenario to study the evolution of social phenotypes.[64] Our data suggest that the ancestral cichlid was a pair-bonding species that provided biparental care. In line with other studies, we found that the evolutionary transition from monogamy and biparental care to non-pair-bonding and maternal care appears to be much more likely than the reverse.[65–67] Therefore, biparental care seems to arise as a consequence of pair-bonding rather than serving as its initial driver.

We hypothesized that the signatures of positive selection observed in the nonapeptide system are associated with the evolution of social phenotypes. As all receptors are expressed in the brain to some extent and represent the primary sites of ligand action, variation in their sequence or expression is likely to influence behavior. To test this, we first examined associations between amino acid variation and the two social traits: pair-bonding and the sex of the caregiver. While no correlation was detected between amino acid variation and caregiver sex, a residue in the extended IL3 of VTR2Aa (position 369) showed evidence of evolving in association with pair-bonding. This SNP results in a valine-to-alanine substitution; given the larger side chain of valine, this change may affect hydrophobic interactions, protein stability or folding, and thus potentially alter receptor function. Notably, this site lies only six amino acids downstream of the position with the strongest signal of positive selection, suggesting that this

region of IL3 may be particularly important for receptor function. Together, these observations indicate the possibility that structural modifications in IL3 of VTR2Aa contribute to variation in receptor signaling linked to pair-bonding. Interestingly, although pair-bonding and caregiving strategy are strongly correlated traits, this amino acid variant was associated only with pair-bonding and not with caregiver sex. This suggests that, despite their ecological linkage, these behaviors may be partly underpinned by distinct genetic mechanisms. In particular, variation in VTR2Aa may influence mate attachment without directly affecting parental care, pointing to partially overlapping but separable molecular pathways.

We assessed the relationship between brain expression levels of nonapeptide system genes and social phenotypes using PGLMMs that accounted for shared evolutionary history, sex, and stable isotope values. Notably, *VTR2Bb* expression was higher in pair-bonding species and in biparental caregivers, and OT expression was elevated in biparental species, supporting an association between components of the oxytocinergic system and cooperative social systems in teleosts. Consistent with prior reports of sex differences in the nonapeptide system,[68,69] we observed marked sex-specific expression patterns. Both ligands, *OT* and *VT*, were more highly expressed in males, whereas VTR2Aa expression was consistently elevated in females. These results indicate sex-biased regulation within both ligand and receptor components of the system, potentially reflecting sex-specific modulation of social and parental behaviors. Stable isotope effects were generally weak, suggesting that ecological variation in diet and habitat (as captured by $\delta^{15}N$ and $\delta^{13}C$) does not strongly drive expression differences in these genes.

Despite the insights provided here, several limitations should be considered. Our gene expression analyses are based on bulk brain RNA sequencing, which averages across heterogeneous cell types and regions and may obscure behaviorally relevant spatial patterns; additionally, intraspecific variation is not captured, likely oversimplifying underlying genetic and expression diversity. Tests of positive selection are further limited by the recent divergence of the Lake Tanganyika radiation, reducing sequence variation and statistical power, and primarily focus on canonical receptor isoforms, potentially overlooking distinct selective pressures on alternative variants. Importantly, associations between genetic variation, gene expression, and social phenotypes are correlative and do not establish causality, and behavioral traits are influenced by multiple genetic and environmental factors only partially accounted for here. Finally, expression was measured under baseline conditions and does not capture dynamic regulation of the nonapeptide system in response to social stimuli, highlighting the need for complementary functional and experimental approaches.

Taken together, our results support a model in which both coding and regulatory components of the nonapeptide system contribute to the diversification of social systems in cichlids. While overall receptor architecture remains highly conserved under strong purifying selection, specific domains involved in intracellular signaling—particularly the extended IL3 of VTR2A receptors—exhibit variability and signatures of adaptive evolution, alongside phenotype-associated sequence variation in VTR2Aa. In parallel, differences in the expression of VTR2Bb and OT are associated with pair-bonding and biparental care, despite generally low expression levels of VTR2 receptors in the brain; importantly, even subtle variation in receptor abundance may have significant effects on neural signaling and behavioral outcomes. This combination of sequence and regulatory variation is consistent with a broader body of work across vertebrates showing that oxytocinergic and vasotocinergic pathways are evolutionarily conserved yet functionally flexible,[22,70] with behavioral diversity arising primarily through modulation of receptor function, expression, and neural localization rather than major structural innovations. In this context, our findings highlight intracellular signaling domains such as IL3 as potential hotspots for functional diversification, alongside regulatory changes shaping gene expression. More broadly, they support the view that repeated transitions in social systems can be achieved through fine-scale tuning of conserved neuroendocrine pathways. Future work integrating functional assays, spatially resolved transcriptomics, and comparative analyses across additional lineages will be essential to determine how variation in receptor signaling and localization translates into behavioral differences, and to place these mechanisms within a general framework for the evolution of sociality in vertebrates.

5 | CONCLUSIONS

Our findings reveal that nonapeptide precursor and receptor genes are subject to strong purifying selection across the Lake Tanganyika cichlid radiation. While positive selection is rare, the few sites under selection are primarily concentrated in the extended IL of the VTR 2A, suggesting potential functional diversification. Additionally, our correlation analyses highlight the involvement of nonapeptide receptors in shaping social phenotypes, reinforcing their role in the evolution of social behavior in cichlids.

AUTHOR CONTRIBUTIONS

Pol Sorigue: Conceptualization; investigation; visualization; writing—original draft. **Walter Salzburger**: Conceptualization; writing—review and editing. **Rui Oliveira**: Conceptualization; supervision; funding acquisition; writing—review and editing.

ACKNOWLEDGMENTS

We thank Susana Varela for her valuable insights and help in the correlated evolution analysis using the BayesTraits program. This study was funded by a grant from the Portuguese Foundation for Science and Technology (Fundação Portuguesa para a Ciência e a Tecnologia—FCT) awarded to RFO (PTDC/BIA-COM/3068/2020). PS is supported by a FCT PhD Fellowship. WS received funding from the Swiss National Science Foundation (208002).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

For transparency, the peer review documents associated with this article are available at https://doi.org/10.1111/jne.70203.

DATA AVAILABILITY STATEMENT

All scripts used in this study are publicly available on GitHub (https://github.com/psorigue/Molecular-evolution-OTVT-in-LT-cichlids), and a permanent archived version is available in Zenodo (https://doi.org/10.5281/zenodo.19077530).

ORCID

Pol Sorigue https://orcid.org/0000-0002-0243-8423
Walter Salzburger https://orcid.org/0000-0002-9988-1674
Rui F. Oliveira https://orcid.org/0000-0003-1528-618X

REFERENCES

1. Balment RJ, Lu W, Weybourne E, Warne JM. Arginine vasotocin a key hormone in fish physiology and behaviour: A review with insights from mammalian models. *Gen Comp Endocrinol*. 2006;147(1):9-16. doi:10.1016/j.ygcen.2005.12.022
2. Jurek B, Neumann ID. The oxytocin receptor: From intracellular signaling to behavior. *Physiol Rev*. 2018;98(3):1805-1908. doi:10.1152/physrev.00031.2017
3. Theofanopoulou C, Gedman G, Cahill JA, Boeckx C, Jarvis ED. Universal nomenclature for oxytocin–vasotocin ligand and receptor families. *Nature*. 2021;592(7856):747-755. doi:10.1038/s41586-020-03040-7
4. Ocampo Daza D, Bergqvist CA, Larhammar D. The Evolution of Oxytocin and Vasotocin Receptor Genes in Jawed Vertebrates: A Clear Case for Gene Duplications Through Ancestral Whole-Genome Duplications. *Front Endocrinol*. 2022;12:1-21. doi:10.3389/fendo.2021.792644
5. Kobayashi A, Hamada M, Yoshida M, et al. Vasopressin-oxytocin–type signaling is ancient and has a conserved water homeostasis role in euryhaline marine planarians. *Sci Adv*. 2022;8(9):eabk0331. doi:10.1126/sciadv.abk0331
6. Oumi T, Ukena K, Matsushima O, et al. Annetocin: An oxytocin-related peptide isolated from the earthworm, Eisenia foetida. *Biochem Biophys Res Commun*. 1994;198(1):393-399. doi:10.1006/bbrc.1994.1055
7. Banerjee P, Joy KP, Chaube R. Structural and functional diversity of nonapeptide hormones from an evolutionary perspective: A review. *Gen Comp Endocrinol*. 2017;241:4-23. doi:10.1016/j.ygcen.2016.04.025
8. Gwee PC, Tay BH, Brenner S, Venkatesh B. Characterization of the neurohypophysial hormone gene loci in elephant shark and the Japanese lamprey: Origin of the vertebrate neurohypophysial hormone genes. *BMC Evol Biol*. 2009;9(1):1-15. doi:10.1186/1471-2148-9-47
9. Sartorius AM, Rokicki J, Birkeland S, et al. An evolutionary timeline of the oxytocin signaling pathway. *Commun Biol*. 2024;7(1):471. doi:10.1038/s42003-024-06094-9
10. Mayasich SA, Clarke BL. The emergence of the vasopressin and oxytocin hormone receptor gene family lineage: Clues from the characterization of vasotocin receptors in the sea lamprey (Petromyzon marinus). *Gen Comp Endocrinol*. 2016;226:88-101. doi:10.1016/j.ygcen.2016.01.001
11. Ivell R, Richter D. Structure and comparison of the oxytocin and vasopressin genes from rat. *Proc Natl Acad Sci USA*. 1984;81(7):2006-2010. doi:10.1073/pnas.81.7.2006
12. Gimpl G, Fahrenholz F. The oxytocin receptor system: Structure, function, and regulation. *Physiol Rev*. 2001;81(2):629-683. doi:10.1152/physrev.2001.81.2.629
13. Grigor'eva ME, Golubeva MG. Oxytocin: Structure, synthesis, receptors, and basic effects. *Neurochem J*. 2010;4(2):75-83. doi:10.1134/S1819712410020017
14. Flores CM, Muñoz D, Soto M, Kausel G, Romero A, Figueroa J. Copeptin, derived from isotocin precursor, is a probable prolactin releasing factor in carp. *Gen Comp Endocrinol*. 2007;150(2):343-354. doi:10.1016/j.ygcen.2006.09.005
15. Barberis C, Mouillac B, Durroux T. Structural bases of vasopressin/oxytocin receptor function. *J Endocrinol*. 1998;156(2):223-229. doi:10.1677/joe.0.1560223
16. Brinton RD, Gonzalez TM, Cheung WS. Vasopressin-induced calcium signaling in cultured hippocampal neurons. *Brain Res*. 1994;661(1–2):174-180. doi:10.1016/0006-8993(94)91194-0
17. Yamaguchi Y, Kaiya H, Konno N, et al. The fifth neurohypophysial hormone receptor is structurally related to the V2-type receptor but functionally similar to V1-type receptors. *Gen Comp Endocrinol*. 2012;178(3):519-528. doi:10.1016/j.ygcen.2012.07.008
18. Wang L, Xu J, Cao S, et al. Cryo-EM structure of the AVP–vasopressin receptor 2–Gs signaling complex. *Cell Res*. 2021;31(8):932-934. doi:10.1038/s41422-021-00483-z
19. Zhang J, Sato M, Duzic E, Kubalak SW, Lanier SM, Webb JG. Adenylyl cyclase isoforms and vasopressin enhancement of agonist- stimulated cAMP in vascular smooth muscle cells. *Am J Physiol Heart Circ Physiol*. 1997;273(2):971-980. doi:10.1152/ajpheart.1997.273.2.h971
20. Liu J, Wess J. Different single receptor domains determine the distinct G protein coupling profiles of members of the vasopressin receptor family. *J Biol Chem*. 1996;271(15):8772-8778. doi:10.1074/jbc.271.15.8772
21. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. 2014;289(6):1045-1060. doi:10.1007/s00438-014-0889-2
22. Carter CS. Neuroendocrine perspectives on social attachment and love. *Psychoneuroendocrinology*. 1998;23(8):779-818. doi:10.1016/S0306-4530(98)00055-9
23. Insel TR, Young LJ. The Neurobiology of Attachment. *Nat Rev Neurosci*. 2001;2:129-136. doi:10.1038/35053579
24. Backström T, Winberg S. Arginine–vasotocin influence on aggressive behavior and dominance in rainbow trout. *Physiol Behav*. 2009;96(3):470-475. doi:10.1016/j.physbeh.2008.11.013
25. Landin J, Hovey D, Xu B, et al. Oxytocin Receptors Regulate Social Preference in Zebrafish. *Sci Rep*. 2020;10(1):5435. doi:10.1038/s41598-020-61073-4
26. Ribeiro D, Nunes AR, Gliksberg M, Anbalagan S, Levkowitz G, Oliveira RF. Oxytocin receptor signalling modulates novelty recognition but not social preference in zebrafish. *J Neuroendocrinol*. 2020;32(4):e12834. doi:10.1111/jne.12834
27. Greenwood AK, Wark AR, Fernald RD, Hofmann HA. Expression of arginine vasotocin in distinct preoptic regions is associated with dominant and subordinate behaviour in an African cichlid fish. *Proc R Soc B: Biol Sci*. 2008;275(1649):2393-2402. doi:10.1098/rspb.2008.0622
28. Saito D, Komatsuda M, Urano A. Functional organization of preoptic vasotocin and isotocin neurons in the brain of rainbow trout: Central and neurohypophysial projections of single neurons. *Neuroscience*. 2004;124(4):973-984. doi:10.1016/j.neuroscience.2003.12.038
29. Kline RJ, O'Connell LA, Hofmann HA, Holt GJ, Khan IA. The distribution of an AVT V1a receptor in the brain of a sex changing fish, Epinephelus adscensionis. *J Chem Neuroanat*. 2011;42(1):72-88. doi:10.1016/j.jchemneu.2011.06.005
30. Mennigen JA, Ramachandran D, Shaw K, Chaube R, Joy KP, Trudeau VL. Reproductive roles of the vasopressin/oxytocin neuropeptide family in teleost fishes. *Front Endocrinol*. 2022;13:1-21. doi:10.3389/fendo.2022.1005863

31. Goodson JL. The vertebrate social behavior network: Evolutionary themes and variations. *Horm Behav*. 2005;48(1 SPEC. ISS):11-22. doi:10.1016/j.yhbeh.2005.02.003
32. Huffman LS, O'Connell LA, Kenkel CD, Kline RJ, Khan IA, Hofmann HA. Distribution of nonapeptide systems in the forebrain of an African cichlid fish, Astatotilapia burtoni. *J Chem Neuroanat*. 2012;44(2):86-97. doi:10.1016/j.jchemneu.2012.05.002
33. Kareklas K, Sorigue P, Oliveira RF. Nonapeptide Evolution and the Regulation of Social Behaviour in Teleost Fish: From Molecules to Sociality. In Grinevich V, Oliveira R, eds. *Evolutionary and Comparative Neuroendocrinology*. Springer, Cham; 2025:441-471. doi:10.1007/978-3-031-80209-6_15
34. O'Connell LA, Hofmann HA. The Vertebrate mesolimbic reward system and social behavior network: A comparative synthesis. *J Comp Neurol*. 2011;519(18):3599-3639. doi:10.1002/cne.22735
35. Lema SC, Washburn EH, Crowley ME, Carvalho PG, Egelston JN, McCormick SD. Evidence for a role of arginine vasotocin receptors in the gill during salinity acclimation by a euryhaline teleost fish. *Am J Physiol Regul Integr Comp Physiol*. 2019;316(6):R735-R750. doi:10.1152/ajpregu.00328.2018
36. Godwin J, Thompson R. Nonapeptides and social behavior in Fishes. *Horm Behav*. 2012;61(3):230-238. doi:10.1016/j.yhbeh.2011.12.016
37. Nowicki JP, Pratchett MS, Walker SPW, Coker DJ, O'Connell LA. Gene expression correlates of social evolution in coral reef butterflyfishes. *Proc R Soc Lond B Biol Sci*. 2020;287(1929):20200239. doi:10.1098/rspb.2020.0239
38. Reddon AR, O'Connor CM, Nesjan E, et al. Isotocin neuronal phenotypes differ among social systems in cichlid fishes. *R Soc Open Sci*. 2017;4(5):170350. doi:10.1098/rsos.170350
39. Ripley JL, Foran CM. Quantification of whole brain arginine vasotocin for two Syngnathus pipefishes: Elevated concentrations correlated with paternal brooding. *Fish Physiol Biochem*. 2010;36(4):867-874. doi:10.1007/s10695-009-9361-3
40. Song Z, Albers HE. Cross-talk among oxytocin and arginine-vasopressin receptors: Relevance for basic and clinical studies of the brain and periphery. *Front Neuroendocrinol*. 2018;51:14-24. doi:10.1016/j.yfrne.2017.10.004
41. Ronco F, Matschiner M, Böhne A, et al. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*. 2021;589(7840):76-81. doi:10.1038/s41586-020-2930-4
42. Salzburger W. Understanding explosive diversification through cichlid fish genomics. *Nat Rev Genet*. 2018;19(11):705-717. doi:10.1038/s41576-018-0043-9
43. Ronco F, Büscher HH, Indermaur A, Salzburger W. The taxonomic diversity of the cichlid fish fauna of ancient Lake Tanganyika, East Africa. *J Great Lakes Res*. 2020;46(5):1067-1078. doi:10.1016/j.jglr.2019.05.009
44. O'Connell LA, Matthews BJ, Hofmann HA. Isotocin regulates paternal care in a monogamous cichlid fish. *Horm Behav*. 2012;61(5):725-733. doi:10.1016/j.yhbeh.2012.03.009
45. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
46. El Taher A, Böhne A, Boileau N, et al. Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat Ecol Evol*. 2021;5(2):243-250. doi:10.1038/s41559-020-01354-3
47. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
48. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-274. doi:10.1093/molbev/msu300
49. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):1-4. doi:10.1093/gigascience/giab008
50. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780. doi:10.1093/molbev/mst010
51. Rajput B, Pruitt KD, Murphy TD. RefSeq curation and annotation of stop codon recoding in vertebrates. *Nucleic Acids Res*. 2019;47(2):594-606. doi:10.1093/nar/gky1234
52. Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;34(WEB. SERV. ISS):609-612. doi:10.1093/nar/gkl315
53. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000;17(1):32-43. doi:10.1093/oxfordjournals.molbev.a026236
54. Kosakovsky Pond SL, Frost SDW. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 2005;22(5):1208-1222. doi:10.1093/molbev/msi105
55. Pond SK, Muse SV. Site-to-site variation of synonymous substitution rates. *Mol Biol Evol*. 2005;22(12):2375-2385. doi:10.1093/molbev/msi232
56. Pagel M. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proc R Soc Lond B Biol Sci*. 1994;255(1342):37-45. doi:10.1098/rspb.1994.0006
57. Pagel M, Meade A. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. *Am Nat*. 2006;167(6):808-825. doi:10.1086/503444
58. Martos-Sitcha JA, Fuentes J, Mancera JM, Martínez-Rodríguez G. Variations in the expression of vasotocin and isotocin receptor genes in the gilthead sea bream Sparus aurata during different osmotic challenges. *Gen Comp Endocrinol*. 2014;197:5-17. doi:10.1016/j.ygcen.2013.11.026
59. Moon TW, Mommsen TP. Vasoactive peptides and phenylephrine actions in isolated teleost hepatocytes. *Am J Physiol Endocrinol Metab*. 1990;259(5):E644-E649. doi:10.1152/ajpendo.1990.259.5.E644
60. Rawat A, Chaube R, Joy KP. In situ localization of vasotocin receptor gene transcripts in the brain-pituitary-gonadal axis of the catfish Heteropneustes fossilis: a morpho-functional study. *Fish Physiol Biochem*. 2019;45(3):885-905. doi:10.1007/s10695-018-0590-1
61. Gonzalez A, Borquez M, Trigo CA, et al. The splice variant of the V2 vasopressin receptor adopts alternative topologies. *Biochemistry*. 2011;50(22):4981-4986. doi:10.1021/bi2001278
62. Postina R, Kojro E, Fahrenholz F. Separate agonist and peptide antagonist binding sites of the oxytocin receptor defined by their transfer into the V2 vasopressin receptor. *J Biol Chem*. 1996;271(49):31593-31601. doi:10.1074/jbc.271.49.31593
63. Waltenspühl Y, Ehrenmann J, Vacca S, Thom C, Medalia O, Plückthun A. Structural basis for the activation and ligand recognition of the human oxytocin receptor. *Nat Commun*. 2022;13(1):1-9. doi:10.1038/s41467-022-31325-0
64. Sefc KM. Mating and Parental Care in Lake Tanganyika's Cichlids. *Int J Evol Biol*. 2011;2011:1-20. doi:10.4061/2011/470875
65. Goodwin NB, Balshine-Earn S, Reynolds JD. Evolutionary transitions in parental care in cichlid fish. *Proc R Soc B: Biol Sci*. 1998;265(1412):2265-2272. doi:10.1098/rspb.1998.0569
66. Iles TD, Holden MJ. Bi-parental mouth brooding in Tilapia galilaea (Pisces, Cichlidae). *J Zool*. 1969;158(3):327-333. doi:10.1111/j.1469-7998.1969.tb02151.x
67. Klett V, Meyer A. What, if Anything, is a Tilapia?—Mitochondrial ND2 Phylogeny of Tilapiines and the Evolution of Parental Care Systems in

the African Cichlid Fishes. *Mol Biol Evol*. 2002;19(6):865-883. doi:10.1093/oxfordjournals.molbev.a004144

68. Albers HE. Species, sex and individual differences in the vasotocin/vasopressin system: Relationship to neurochemical signaling in the social behavior neural network. *Front Neuroendocrinol*. 2015;36:49-71. doi:10.1016/j.yfrne.2014.07.001
69. Dumais KM, Veenema AH. Vasopressin and oxytocin receptor systems in the brain: Sex differences and sex-specific regulation of social behavior. *Front Neuroendocrinol*. 2016;40:1-23. doi:10.1016/j.yfrne.2015.04.003
70. Goodson JL, Kingsbury MA. Nonapeptides and the evolution of social group sizes in birds. *Front Neuroanat*. 2011;5(13):9. doi:10.3389/fnana.2011.00013

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sorigue P, Salzburger W, Oliveira RF. Nonapeptide molecular evolution during the adaptive radiation of Tanganyika cichlids. *J Neuroendocrinol*. 2026;38(6):e70203. doi:10.1111/jne.70203