Enrichment of oxytocin pathway genes in human evolution

Alina M. Sartorius[1,2], Constantina Theofanopoulou[8], Francesco Bettella[1], Jaroslav Rokicki[1,2], Claudia Barth[1], Ann-Marie G. de Lange[1,2,4], Marit Haram[1], Alexey Shadrin[1], Dennis van der Meer[1,5], Adriano Winterton[1], Nils Eiel Steen[1], Emanuel Schwarz[6], Dan Stein[7], Lars T. Westlye[1,2,3], Ole A. Andreassen[1,3], Daniel S. Quintana[1,2,3]*

***For correspondence:** daniel.quintana@medisin.uio.no (DSQ)

[1]NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo, and Oslo University Hospital, Oslo, Norway; [2]Department of Psychology, University of Oslo, Oslo, Norway; [3]KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway; [4]Department of Psychiatry, University of Oxford, Oxford, UK; [5]School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands; [6]Central Institute of Mental Health, Department of Psychiatry and Psychotherapy, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [7]Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa; [8]Jarvis Lab, Laboratory of Neurogenetics of Language, Rockefeller University/HHMI, New York, New York, United States of America

Abstract Oxytocin is a hormone and neuromodulator that is associated with several behavioral and physiological processes across widely diverged mammalian species. Ancestral versions of oxytocin have been identified in species that share ancient common ancestors with humans, yet the evolutionary timeline of the emergence of the oxytocin system signaling has yet to be systematically delineated and it is unclear if the oxytocin signaling system has undergone change in more recent evolutionary history. By using phylostratigraphy, which identifies the phylogenetic stage that a gene first emerged, we show that most genes in the oxytocin signaling pathway are ancient, as they first appeared during genesis of cellular organisms. The evolutionary age of the oxytocin signalling transcriptome across 16 brain regions is relatively stable across ontogeny, however there is modestly higher expression of younger genes prenatally. Ancient oxytocin signaling genes are highly expressed in contracting organs (e.g., heart and muscle) while more modern genes are highly expressed in the brain. Comparing the oxytocin pathway gene dN/dS ratios of eight primate species, including humans, with their common primate ancestor suggests that genomic sequences of oxytocin signaling pathway genes were under purifying selection but less so in the recent primate species. Altogether, our results suggest that oxytocin signaling initially emerged as a system supporting muscle contraction but more recently developed a role in behavior and cognition.

Introduction

The oxytocin system is implicated in several physiological functions, such as reproduction, cardiovascular homeostasis, metabolism, and bone regeneration (***Jurek and Neumann, 2018***; ***Winterton et al., 2020***). More recently, oxytocin has been linked to social behavior and cognition (***Guastella***

and Hickie, 2016). As social dysfunction is a key feature of several mental illnesses (e.g., bipolar disorder, schizophrenia, and autism), researchers have investigated oxytocin's potential as a therapeutic to help alleviate social difficulties. Early work exploring the role of the oxytocin system in social behavior investigated oxytocin receptor (*OXTR*) gene knockout animal models, which revealed the importance of the oxytocin system in mammalian behavior. For instance, *OXTR* knockout mice exhibit deficits in social recognition and memory, which can be rescued by central oxytocin infusion (*Winslow and Insel, 2002*). Divergent central *OXTR* expression patterns in the pair-bonding prairie vole compared to the closely-related monogamous meadow vole also underscore the role of the oxytocin system in social behavior (*Insel and Shapiro, 1992*). One hundred and fifty four genes currently support oxytocin signaling in humans, but the most intensively studied genes in terms of human behavior are *OXTR*, encoding for the oxytocin receptor, the structural gene for oxytocin (*OXT*), and *CD38*, which regulates oxytocin secretion (*Jurek and Neumann, 2018*).

Oxytocin and oxytocin-like signaling can be observed in a wide range of mammals, fish, birds, reptiles, mollusks, insects, and tunicates, which reflects the long evolutionary history of the oxytocin system (*Feldman et al., 2016*). There is also preliminary evidence for the peptides from the oxytocin-vasotocin family (*Theofanopoulou et al., 2021*) playing a role in as ancient phyla as cnidaria, but evidence in this area is very sparse, and therefore needs further investigation (**Takahashi, 2020**). While oxytocin-like signaling has been associated with basic social behaviors across these species, complex social behaviors that reflect a sophisticated understanding of the goals and intentions of conspecifics—from dyadic romantic relationships to the cooperation of entire communities—is almost exclusively a human trait (*Warneken and Tomasello, 2006*). The central oxytocin receptor system is expressed in different brain regions to promote species-specific social behaviors. For example, in rodents who preferably rely on olfactory cues, there is greater *OXTR* expression in the olfactory bulb (*Vaccari et al., 1998*). Variation in the *OXTR* gene also predicts central expression and social attachment in monogamous prairie voles (*King et al., 2015*). Differences in oxytocin system function have also been observed between non-human primates, which may underlie unique social behavior traits in tamarins and marmosets (*Vargas-Pinilla et al., 2015*), for instance. In human adults, there is increased *OXTR* expression in brain regions associated with the processing of anticipatory, appetitive, and aversive cognitive states, such as the amygdala and pallidum (*Quintana et al., 2019*). Moreover, *OXTR* concentrations in humans peak during early childhood (*Rokicki et al., 2021*), which is a critical period for social learning (*Nielsen, 2012*). The involvement of the neurohypophysial hormone system in a multitude of bio-physiological processes across species has been demonstrated in several studies. For instance, it plays a role in parturition in annelids and different mammals including humans and possibly horses (**Fujino et al., 1999**; **Gram et al., 2014**; **Kim et al., 2017**; **Rapacz-Leonard et al., 2020**), and has been implicated in the regulation of energy balance and metabolism in rodents and primates (**Blevins et al., 2015**; **Camerino, 2009**; **Yuki et al., 2008**).

As close as primate species are related to each other genetically, and consequently resemble each other phenotypically, they still exhibit remarkable socio-behavioral, locomotory, and anatomical diversity. Hominoids distinguish themselves from new world and old world monkeys for example through larger brains, the absence of philopatry in females and a tail, which has effects on locomotion, and longer gestation periods (**Feagle, 2013**). There is also considerable archaeological and palaeontological evidence that our ancestors began living together in large complex groups during the Pleistocene period (*Hublin et al., 1996*), just after divergence from Neanderthals (**Gómez-Robles, 2019**). Indeed, the evolutionary success of modern humans has also been attributed to social cooperation in complex group structures (*Hill et al., 2011*; *Powell et al., 2009*). It is unclear, however, whether the oxytocin system adapted along these more recent social challenges in response to selection pressures faced by modern human ancestors and whether it underwent adaptive changes with the divergence of preceding primate lineages. Moreover, despite the seemingly increasing consensus about the oxytocin system being an evolutionary conserved system (**Carter, 2014**; (*Feldman et al., 2016*); **Knobloch & Grinevich, 2014**), it hasn't been quantified when the

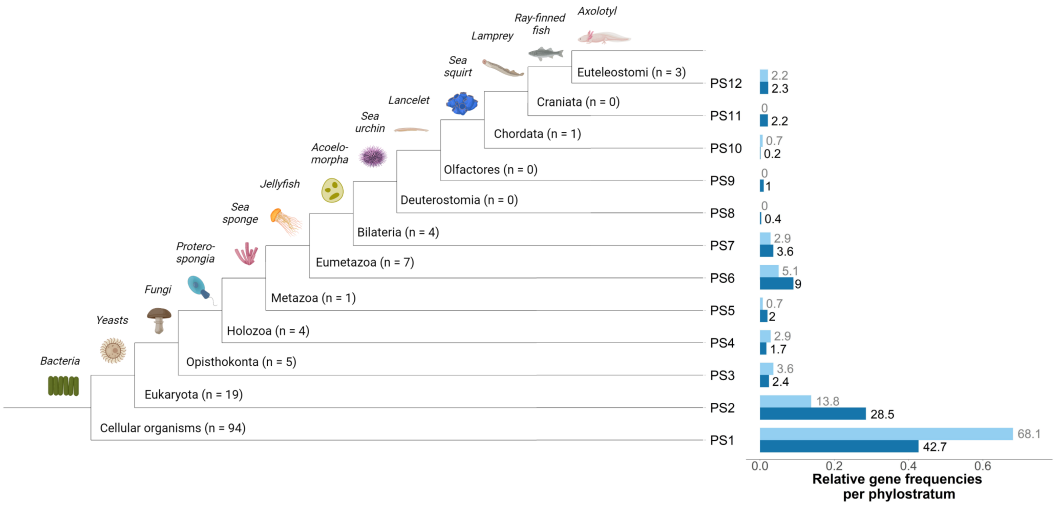

Figure 1. The evolutionary age of the oxytocin signaling pathway. The phylogenetic stages that the ancestors of genes in the oxytocin signaling pathway first appeared, lower numbers indicate earlier phylostrata. A representative species is shown for each stage. The number of genes from the oxytocin pathway per stage is shown in parentheses. The relative gene frequencies in percent (represented on the x-axis) per phylostratum (represented on the y-axis) from the oxytocin signaling pathway in light blue and from entirety of protein-coding genes (n = 18844) in dark blue are shown on the right. Phylostrata 13-19 are not displayed since no oxytocin pathway gene is assigned to any of these stages, however the remaining 7.2% of protein-coding genes emerged during these phylostrata.

oxytocin signalling system as a whole emerged during evolution. Identifying the eras during which genes of a subset of interest originated using phylostratigraphy can put a tangible number on the age of the oxytocin system. It can also help clarify the functions of the human oxytocin system by testing whether these appearance of genes coincide with specific evolutionary or developmental advancements. Furthermore, while the three oxytocin candidate genes *OXT*, *OXTR*, and *CD38* have received considerable research attention, less is known about the entirety of genes that make up the oxytocin signaling pathway. The oxytocin system is complex, influencing various physiological, biological, cognitive, and social processes. Therefore, considering the whole set of genes that support the oxytocin signaling pathway, next to the three oxytocin candidate genes, will help unravel the function and the evolutionary history of the oxytocin system.

The goal of this study was to systematically explore the evolutionary age of the 154 genes that constitute the oxytocin signaling pathway with phylostratigraphy, and to connect the age of genes with evolutionary advances and functional properties of this system. We also assessed the age of the transcriptome of the oxytocin pathway genes in the human brain across ontogeny by computing transcriptomic age indices, and whether genetic variation in the oxytocin system pathway has undergone recent natural selection by evaluating dN/dS ratios in humans, other *Hominini* and primate ancestors. Altogether, our goal was to provide a concise and annotated evolutionary timeline of the oxytocin system that could substantiate and expand the current view of the ancient oxytocin system.

Results

The evolutionary age of genes in the oxytocin signaling pathway

Phylostratigraphy is a method that can reveal the evolutionary origin of genes by identifying which branch of the phylogenetic tree (***Figure 1***) the original ancestor of a gene first appeared via the comparison of genomes across species (***Domazet-Lošo et al., 2007***). Performing phylostratigraphy on the oxytocin signaling pathway gene set revealed that most genes (118 out of 138) in the pathway

are evolutionary ancient, as they first appeared during the first three ancient phylostrata, which include genes found in unicellular organisms like bacteria and yeasts (***Figure 1***). While the majority of all protein-coding genes emerged during the early phylostrata of cellular organisms (1), eukaryota (2), and opisthokonta (3), oxytocin pathway genes still are relatively abundant in the first and third phylostratum (68.1% vs. 40.2% and 3.6% vs. 2.4%; ***Figure 1*** right). Among the younger phylostrata, most oxytocin genes originated during phylostratum 6 and 7 (5.1% and 2.9%), as did the majority of human protein-coding genes. Of note, the first ancestor of the oxytocin receptor and *CD38* genes emerged during the Eumetazoa phylostratum, during which basic nervous systems and organ systems appeared (***Kelava et al., 2015***). The oxytocin gene's ancestor first appeared during the Biletaria stage, which is characterized by volitional and centralized motor control and a more complex, rudimentarily centralized nervous systems (***Nielsen, 2008***). The observed genes-per-phylostrata distribution was also tested for enrichment of phylostrata to examine whether any of the phylostrata were statistically meaningful(ly) over- or under-represented within the oxytocin pathway gene set. Indeed the first phylostratum, during which cellular organisms developed, was significantly enriched in the gene set of interest; the phylostratum is present at a higher rate in the oxytocin pathway gene set compared to the full set of protein-coding genes ("background gene set"; *pFDR* = 9.61e-09; ***Figure 2***). On the contrary, phylostratum 2 (eukaryota development) was substantially under-represented in the oxytocin pathway gene set in relation to the background gene set, comparatively less oxytocin pathway genes could be categorized in this phylostratum (*pFDR* = 5.56e-04). The OT gene set also visually appeared to be enriched for phylostratum 5 and 10 (metazoa and chordata, respectively), however these did not pass the statistical significance threshold (both *p*'s > 0.05), possibly due to too few genes in the respective phylostratum (n = 1 in each).

Phylostratigraphy was extended with evolutionary transcriptomics, which can quantify the evolutionary age of a transcriptome across ontogeny by estimating the relative contributions of ancient and modern genes to a transcriptome at an ontogenetic stage. Here we used previously described phylogenetic stages (***Domazet-Lošo and Tautz, 2008***), which are also illustrated in ***Figure 1***. As oxytocin pathway gene expression patterns shift across ontogeny in both animals (***Vaidyanathan and Hammock, 2017***) and humans (***Rokicki et al., 2021***), RNA sequencing data from five ontogenetic stages (prenatal, infant, child, adolescent, adult) across sixteen brain was retrieved from the BrainSpan database to assess putative differences in the evolutionary conservation of these transcriptomes. For each brain region, a transcriptional age index (TAI) was calculated for each ontogenetic stage by calculating the average age of genes that contribute to the transcriptome (***Domazet-Lošo and Tautz, 2010***). To compute the TAI, the expression level for each each gene is multiplied by its gene age (i.e., the phylostratum), and then the values for each gene in a gene set are averaged. Therefore, lower values represent an older transcriptome age. In our gene set sample, because phylostrata 1 to 7, 10, and 12 are represented, the TAI could in theory assume a value anywhere between 1 – exclusively genes from phylostratum 1 are expressed – and 12 – only genes from phylostratum 12 are expressed (**Šestak et al., 2013**). The transcriptional age of the oxytocin pathway genes ranges from 1.43 to 1.72. It is at a maximum during the prenatal period, plunges during infancy and childhood, rises to a small peak around adolescence and decreases to a low point at adulthood for the majority of the brain regions (62.5%) (all *p*-values > 0.05; ***Figure 2***). The TAI trajectory in the striatum (STR) deviates from this pattern: As opposed to the majority pattern, it peaks again during childhood, drops around adolescence and remains stable during adulthood. The trajectories in the amygdala (AMY), primary motor cortex (M1C) and inferior parietal cortex (IPC) are characterized by a decline at adolescence, and trajectories of the mediodorsal nucleus of the thalamus (MD) and primary visual cortex (V1C) by a moderate increase at childhood. Importantly, there was no statistical difference in transcriptional age between ontogenetic stages in any brain region (all *p*-values > 0.05; ***Figure 2***), suggesting that the age of the mRNA profile of the oxytocin pathway genes are stable across ontogeny.

Genes in the oxytocin pathway were classified as ancient and modern based on their timepoint

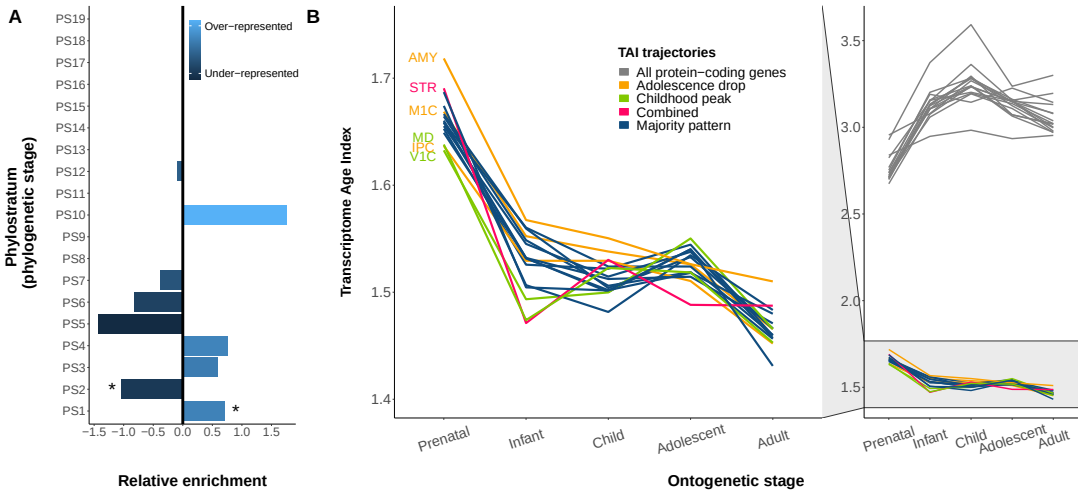

Figure 2. Enrichment of phylostrata and TAI. (A) Enrichment (over- and under-representation, x-axis) of phylostrata ("PS", y-axis) within the oxytocin signaling pathway gene set. Phylostratum 1 (cellular organsisms) is statistically significantly over-represented in the oxytocin gene set, phylostratum 2 (eukaryota) significantly under-represented (*p*-values < 5e-03). (B) The transcriptional age indices of the oxytocin signaling pathway genes (colorful lines) compared to the transcriptional age indices of all protein-coding genes in 16 brain region in the human on the right, side part of the figure B. The indices of the oxytocin gene set are magnified on the left, main part of the figure B. In both the main and the side part, the y-axis indicates the TAI, the x-axis the five ontogenetic stages.

of origin. 118 genes were located in the three earliest phylostrata around the genesis of living organisms (***Figure 1***) and are thus ancient, and 20 genes were located in the more 'recent' phylostrata (4 - 12; no genes in phylostrata 8, 9, and 11, and 13-19) and hence termed 'modern'. These ancient and modern gene sets were submitted to the online tool FUMA (***Watanabe et al., 2017***), which integrates gene expression data from the GTEx database and GWAS data, to better understand the functional relevance of a given gene set. This revealed that genes from the ancient oxytocin gene set are enriched in systolic blood pressure (p = 8.18e-3), four body mass and obesity related (p = 3.58e-2 - p = 1.53e-2), two tooth decay/dentures (both p = 8.18e-3), one hand grip strength (p = 3.14e-2) and a pubertal anthropometrics (p = 4.56e-2) GWAS. Across all gene sets, the ancient OT genes were most strongly enriched in the KEGG Vascular Smooth Muscle Contraction gene set (p = 2.73e-70; all p-values FDR corrected). Out of thirty tissue types across the body, the ancient oxytocin gene set was up-regulated in arterial, skeletal muscle, bladder, and heart tissue (Bonferroni corrected p-values 2.44e-2 - 2.79e-7; ***Figure 3***). The modern oxytocin pathway gene set was most significantly enriched in the Reactome gene set for the plateau phase of the cardiac action potential (p = 1.23e-18); this phase is responsible for sustaining muscle contraction (**Grant, 2009**). The modern gene set was further solely up-regulated in brain tissue (Bonferroni corrected p = 1.39e-2; ***Figure 3***).

Given the latter result, expression of the modern oxytocin pathway gene set in the brain was investigated. Between-brain region comparisons of the expression patterns of the modern gene set were calculated for 42 brain regions (left-hemisphere, cortical and subcortical regions). The mRNA intensities of the oxytocin gene set in each of the 42 brain regions were compared against the average mRNA intensity of the oxytocin gene set across all brain regions (whole-brain). This yielded statistically significant mean differences in three sub-cortical brain regions (thalamus proper: p = 4.97e-3, d = 4.41, t = -10.79; brain stem: p = 2.27e-2, d = 2.76, t = -6.75; pallidum: p = 2.64e-2, d = 2.44, t = -5.96; all p-values FDR corrected), indicating that the more modern oxytocin pathway gene set is down-regulated in certain sub-cortical, central structures (***Figure 4***).

In addition, within-brain region comparisons were performed (i.e., region-specific expression profiles of the modern OT pathway gene set versus region-specific average expression profiles of

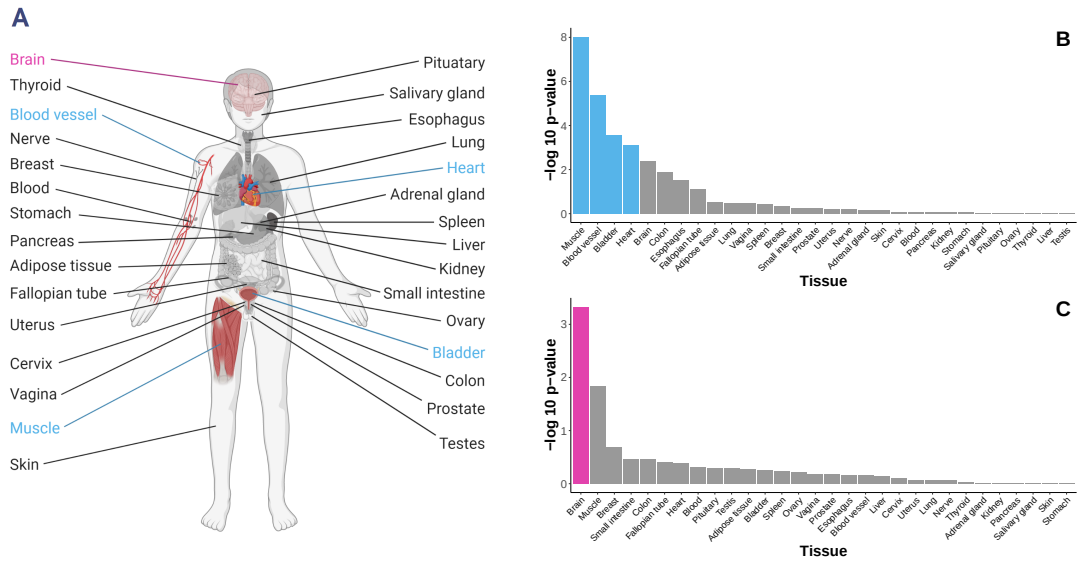

Figure 3. (A) Differential expression of genes from the oxytocin signaling pathway in 30 tissue types from the the GTEx dataset. Significantly enriched tissues are coloured. (B) There is an increased expression of the ancient gene set in muscle, blood vessel, bladder, and heart tissue. (C) Expression of the more modern genes is enriched in the brain. -log 10 p-values represent the probability of hypergeometric tests.

all protein-coding genes). Modern oxytocin genes were up-regulated in the isthmus of the cingulate cortex (*pFDR* = 4.5e-2, *d* = 2.09), the transverse temporal cortex (*pFDR* = 4.5e-2, *d* = 1.94), and in the entorhinal cortex (*pFDR* = 4.5e-2, *d* = 1.94), and down-regulated in the thalamus proper (*pFDR* = 1.56e-2, *d* = 3.47) and pallidum (*pFDR* = 4.5e-2, *d* = 1.92; ***Figure 5***).

Patterns of natural selection across primate species

Using the GenEvo online tool and the underlying calculation algorithm implemented by **Dumas et al., 2021**, the dN/dS ratios of 107 oxytocin signaling pathway genes for eight primates species, including the modern human and two further *Hominini*, in relation to their common primate ancestor, were obtained. The dN/dS or Ka/Ks ratio (ω) is among the most widely applied measures to quantify selective pressure on orthologous gene sequences (**Kryazhimskiy and Plotkin, 2008**). It estimates the proportion of non-synonymous substitutions to synonymous substitutions at different sites within a genomic region between two divergent lineages, or, simplified, the rate of substitutions at a nucleotide site, and yields a value ranging from 0 to theoretically infinity. The resulting parameter is usually interpreted as follows: $\omega < 0$ indicates negative or purifying selection, $\omega = 1$ signifies neutral selection, and $\omega > 1$ points to positive or beneficial selection. The dN/dS ratio distribution for our gene set of interest revealed that the vast majority of genes was under selective constraint in all eight primate species genomes in comparison to the genome of their shared primate ancestor (mean dN/dS ratios per species range 0.09 - 0.24), indicating strong conservation of these genes throughout the evolution of the analyzed primate species.

A slight but distinct shift in ω distributions towards $\omega = 1$ occurred in the *Hominini* and *Panini* sister tribes which comprises *Homo sapiens* (humans), *Homo neanderthalensis* (neanderthals (altaï)), *Homo denisova* (Denisovans), and *Pan troglodytes* (common chimpanzee). The gene that is involved in oxytocin secretion, *CD38*, stands out in the three most distant primate species, *Callithrix jacchus* (common marmoset), *Macaca mulatta* (rhesus macaque), and *Pongo abelii* (sumatran orangutan), as it's ω value is noticeably higher than in the other genes and species and trends towards a neutral selection ($\omega_{CD38callithrix} \approx 0.79$, $\omega_{CD38macaca} \approx 0.91$, $\omega_{CD38pongo} \approx 0.65$). Lastly, we identified three genes that were under positive selection in relation to a shared ancestor: *PPP1R12A* in the common chimpanzee ($\omega_{PPP1R12A} \approx 1.38$) and *MYLK4* and *CALML4* in the altaï neanderthal ($\omega_{MYLK4} \approx 1.00$ and

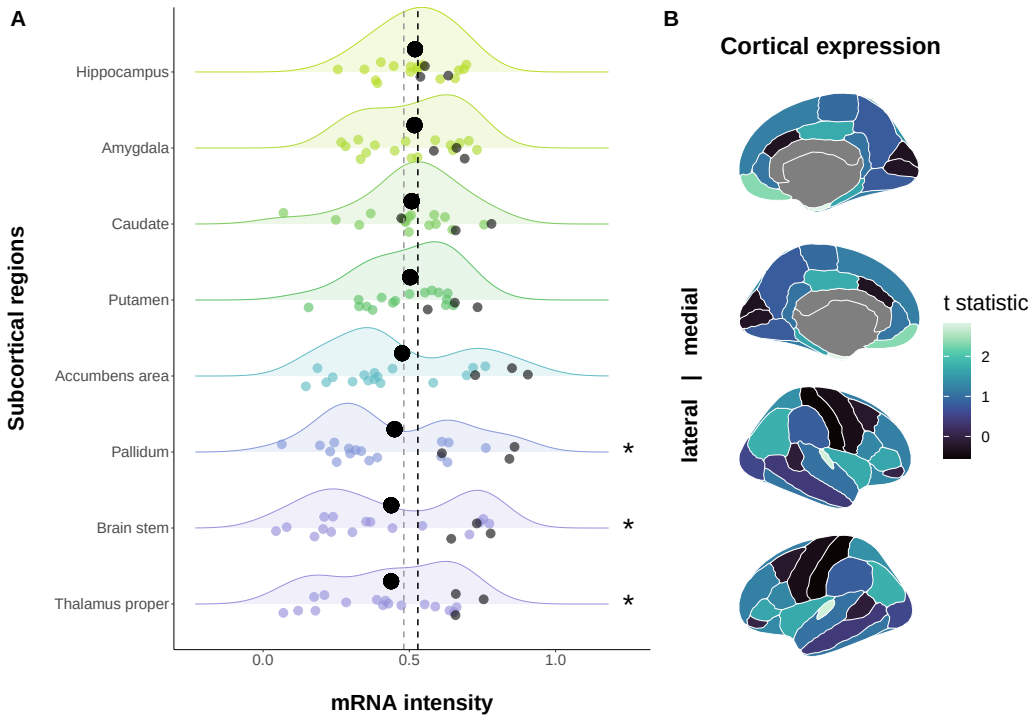

Figure 4. Expression of the modern oxytocin pathway genes across the brain. (A) Mean expression of the modern oxytocin gene set in 8 sub-cortical regions (bold black dots) compared to the average expression of the modern oxytocin gene set across the whole brain (vertical black dashed line; for reference the vertical grey dashed line represents mean expression across sub-cortical regions). Jittered colored dots represent the mRNA intensity of each modern oxytocin gene, the three grey dots in each distribution represent mRNA intensity of *OXT*, *OXTR*, and *CD38*. * = *p*FDR < 0.05. (B) Atlas representation of the *t*-values in 34 cortical brain regions (unilateral left). High *t*-statistics can be observed in the entorhinal cortex, transverse temporal cortex, temporal pole, and medial orbito-frontal cortex, low values appear in the post-central gyrus, pre-central gyrus, cuneus, and the caudal ACC. The raw data underlying these plots is presented in supplementary materials 2 & 3.

225 $\omega_{CALML4} \approx 1.12$, respectively).

226 Discussion

227 Unlike bones and artifacts, human behavior and biological processes leave very little historical re-
228 mains. However, comparative genetics can offer a window into the behavior and physiology of our
229 ancestors, consequently providing information on the origins of human behavior and biological
230 processes. Here, our analyses indicate that most genes from the oxytocin pathway are ancient
231 and strongly conserved throughout evolution. Over 85% of oxytocin pathway gene homologues
232 and precursors emerged during the first three phylostrata, commonly agreed upon as 'ancient' or
233 'old' phylostrata, and within this sub-genset almost 80% in phylostratum 1. Accordingly, oxytocin
234 pathway genes were significantly enriched in the first phylogenetic stage compared to the entirety
235 of protein-coding genes, despite most protein-coding genes stemming from this phylostratum. Ba-
236 sic unicellular organisms, eukaryota, and opisthokonta are the lifeforms originating from these
237 phylostrata and comprise organisms like bacteria and archaea, yeasts, and other fungi (**Cooper,**
238 **2000**). While primitive forms of social interaction among bacteria have been proposed and may
239 be possible (**Madsen et al., 2018**), they are not as complex as in higher developed organisms. It
240 is more likely that basic physiological and biological mechanisms like cell division, cell structure,
241 and energy metabolism played the major role in the early stages of evolution and then evolved
242 into multi-faceted physiological functions (e.g., gut digestion, energy levels) along with increasing
243 complexity of organisms. Indeed, we found that ancient oxytocin pathway genes are primarily im-

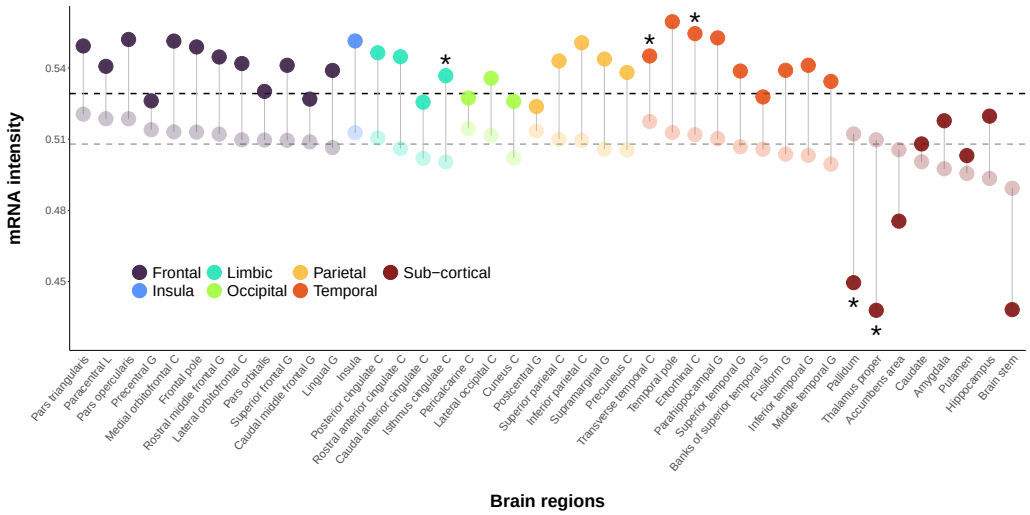

Figure 5. Within-brain-region expression. Compared to the average expression of all protein-coding genes in each brain region (light coloured dots), the average expression of the modern oxytocin pathway genes (dark/bright coloured dots) are enriched in the cerebral structures isthmus of the cingulate cortex, transverse temporal cortex, and entorhinal cortex, and decreased in the sub-cortical structures thalamus proper and pallidum (* = *p*FDR < 0.05). The y-axis indicates expression intensity, 34 cortical and 8 sub-cortical brain regions are represented on the x-axis. C = Cortex, G = Gyrus, L = Lobule, S = Sulcus.

plicated in physiological/biological mechanisms like smooth muscle contraction, which is in line with the hormone's role in for instance uterine, heart, and mammary cell contraction (**Gutkowska & Jankowski, 2012**; **Tom & Assinder, 2010**; **Uvnäs-Moberg et al., 2019**), metabolism and feeding related phenotypes, or blood pressure, but also hand grip strength, which can predict cardiovascular and all-cause mortality (**Kim et al., 2017**; **Lawman et al., 2016**). Associations of oxytocin with hand grip strength, cardiovascular regulation, and body weight have also previously been reported (e.g., **Blevins et al., 2015**; **Gutkowska & Jankowski, 2012**; **Tom & Assinder, 2010**; **Winterton et al., 2021**). Accordingly, the genes tend to be highly expressed in muscle tissue, blood vessels, the bladder, and the heart. One conceivable interpretation therefore could be that ancient oxytocin pathway genes were initially involved in physiological functions suited for early evolutionary stages, and were 'refined' according to evolutionary changes and more complicated biological systems of divergent phyla.

Considerably less genes of the oxytocin pathway emerged during the more recent phylostrata 4 - 12. More than half of the modern OT signaling genes (55%) appear to have originated from the sixth and seventh phylostratum, of note, these phylogenetic stages denote a peak of gene origins among protein-coding genes in general. Phylostratum 6 and 7 are the periods for which it is assumed that an established, primitive central nervous system first appeared (discussed below). Coinciding with this evolutionary developmental stage, we found that expression of more modern OT pathway genes was strongly and solely enriched in brain tissue compared to 29 other tissue samples. Given this result, we then focused in detail on isolated cerebral gene expression, which revealed that modern oxytocin signaling pathway genes are significantly less expressed in the sub-cortical regions pallidum, thalamus proper, and brain stem compared to the averaged expression of the gene set across brain regions, while genes were equally expressed in all remaining brain regions. At a first glance, this may appear contradictory to other studies having reported an increased *OXTR* expression in the pallidum (***Quintana et al., 2019***). *OXTR*, *CD38*, and *OXT*, however, were among the top genes with the highest expression values in all eight sub-cortical structures, most pronounced in the thalamus proper, pallidum, accumbens area, and putamen (***Figure 4***). Interestingly, four out of eight sub-cortical gene expression distributions were bi-modal, among which the three significant ones, indicating that genes in the modern oxytocin pathway either tend

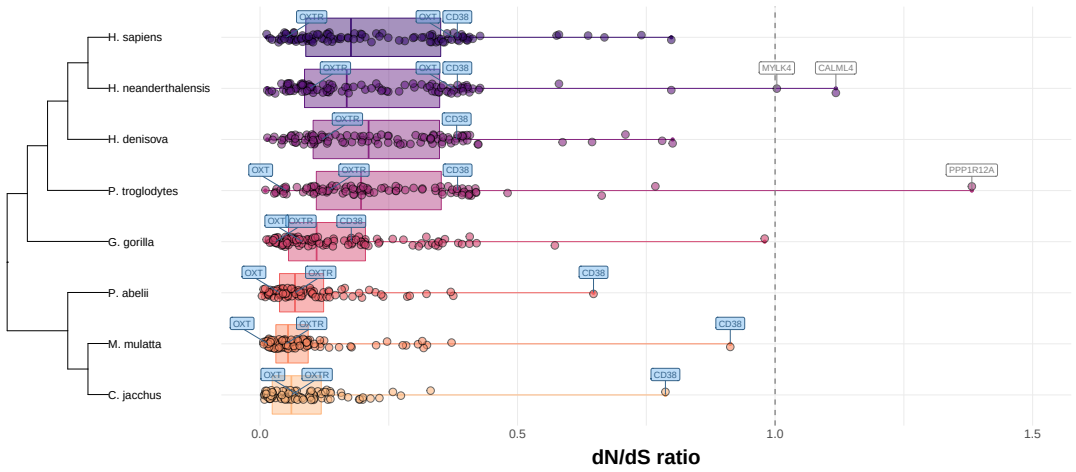

Figure 6. dN/dS ratio in eight primate species. Distribution of the dN/dS ratio of 107 oxytocin pathway genes in different primate species versus one common primate ancestor. Three genes, *PPP1R12A* in P. troglodytes (common chimpanzee), and *MYLK4* and *CALML4* in H. neanderthalensis (Altaï neanderthal), with dN/dS ratio > 1 can be observed. *CD38*, involved in oxytocin secretion, has markedly higher ratios in the first three primate species compared to the other five species. Average ratios are higher in the four most recent species. H. sapiens = *Homo sapiens*, H. neanderthalensis = *Homo neanderthalensis*, H. denisova = *Homo denisova*, P. troglodytes = *Pan troglodytes*, G. gorilla = *Gorilla gorilla*, P. abelii = *Pongo abelii*, M. mulatta = *Macaca mulatta*, C. jacchus = *Callithrix jacchus*.

to the upper or lower extremes of expression intensities in half of the sub-cortical structures, and that the sub-cortical expression signatures of genes within the modern subset are rather heterogeneous. Lower mRNA intensities of the modern gene set were also found in similar sub-cortical regions (pallidum, thalamus proper) when compared to the expression of all protein-coding genes within a given region, and higher mRNA intensities in three cortical areas (isthmus of the cingulate cortex, transverse temporal cortex, entorhinal cortex). The entorhinal cortex has been previously linked to the neuropeptide oxytocin in association with the olfactory system (**Oettl & Kelsch, 2017**) and has been described as either a dense binding site for vasopressin or as a region with high density of respective receptors in mice (**Freeman et al., 2015**; **Gould & Zingg, 2003**, **Young et al., 1999**). The connection between the other enriched brain regions is less obvious. The isthmus of the cingulate cortex has been sporadically implicated in SCZ/psychosis (**Wei et al., 2020**), Alzheimer's disease/cognitive impairments (**Gould & Zingg, 2003**, **Yang et al., 2019**), and depression in elderly (**McLaren et al., 2016**; **Lim et al., 2012**), but it's role in the oxytocin system as well as the role of transverse temporal cortex remains an open question.

The oxytocin hallmark genes *OXTR* and *CD38*, and *OXT* are assigned to more recent phylostrata 6 and 7, respectively (**supplementary materials 1**). Phylostratum 6 marks the origin of eumetazoa (multicellular animal phyla, e.g., cnidarians) and with them a distinct, but not yet centralized nervous system and a separated locomotory ectoderm, about 630 - 850 million years ago (**Dohrmann & Wöhrheide, 2017**; (***Kelava et al., 2015***); (***Nielsen, 2008***); **Peterson et al., 2005**). As briefly mentioned before, it seems possible that cnidaria show oxytocin/vasopressin-like signaling, but there is currently not enough clear evidence for any further assumptions. It is thought that bilateria (phylostratum 7), which includes worms and ancestors of modern vertebrates, emerged about 540 - 750 million years ago (**Dohrmann & Wöhrheide, 2017**; **Malakhov, 2010**). With bilaterian animals, evolution introduced a vastly diverse group of phyla lineages and, presumably, novel higher-order features like bilateral symmetry, triploblasty (instead of diploblasty), a through-gut, circulatory system, and a rudimentary ancestor of the centralized nervous system (**Arendt et al., 2008**; **Denes et al., 2007**, **Extavour, 2007**; **Finnerty, 2003**). Leeches (phylum of annelida), whose ancestors presumably stem from the bilaterian age, show effects of the oxytocin system in reproductive and

digestive processes (**Feldman et al., 2016**; **Fujino et al., 1999**). Interestingly, the remaining genes of the seventh phylogenetic stage are sub-units γ 2, 3, and 7 of the voltage-gated calcium channel γ sub-units (4, 5 and 8 were not available in the expression data set) which are highly, partly exclusively, expressed in the brain (**Arikkath & Campbell, 2003**; **Chen et al., 2007**).

The youngest identified genes in the OT pathway (*CACNG6*, *CACNG1*, *NPPA*) appeared with the original ancestor of bony vertebrates (i.e., the Euteleostomi clade) in the 12th phylostratum, approximately 420 million years ago (**Zhu et al., 2009**). *CACNG6* and *CACNG1* belong to a shared cluster within the family of voltage-gated calcium channel γ sub-units. The two genes are predominantly expressed in muscle tissue, with sub-unit γ 6 being expressed in the brain as well; it is unclear whether sub-unit γ 1 RNA is present in cerebral tissue (**Burgess et al., 2001**; **Chen et al., 2007**). Together with other previously mentioned voltage-gated calcium channel sub-units, *CACNG6* has been associated with schizophrenia risk (**Guan et al., 2016**). *NPPA* encodes for the artrial natriuretic peptide precursor, primarily synthesized in the heart, which is implicated in electrolyte homeostasis and extracellular fluid control, and as a consequence thereof in blood pressure/volume and energy metabolism regulation, and cardiac disease (**Houweling et al., 2005**; **Song et al., 2015**). Lastly, to interpret these results adequately, it is necessary to mention that the phylostratigraphy approach we applied, which has been introduced by **Domazet-Lošo & Tautz (2008)**, has been subject to criticism for underestimating gene age inferences at worst leading to biased assumptions about gene origins (**Moyers & Zhang, 2015**). This critique has been refuted, and it was demonstrated that the presumable gene age assignment error rate between 5 to 15% does not affect the deduced presumptions about phylostratigraphic patterns, that is, they remain stable (**Domazet-Lošo et al., 2017**). While the current phylostratigraphy approach has it's limitations and may need further refinement, it is a robust method to estimate gene age and infer assumptions about gene emergence.

The pattern of evolutionary conservation of genes of the oxytocin signaling pathway is also reflected in the transcriptomic age of the gene set. We observed markedly lower transcriptome age indices for all ontogenetic stages across brain regions in the OT pathway gene set compared to the TAIs of all human protein-coding genes. The transcriptome of the OT signaling pathway gene subset appears to be considerably older than the transcriptome of all protein-coding genes. This makes sense considering that the majority of genes that constitute the gene set emerged during phylostratum 1. Values fluctuate slightly in some areas (e.g., STR, V1C, AMY). All fluctuations across brain regions and ontogeny were statistically not significant, which suggests that throughout developmental stages, a similar ratio of expression of old versus modern oxytocin genes is maintained. This is interesting because the transcriptome age signature of all protein-coding genes often fluctuates substantially over the life-span, as noticeable in humans with lowest TAIs before birth and highest TAI values during childhood (***Figure 2***). Irrespective of the lack of statistical significance, it should be briefly mentioned that the course of the majority of oxytocin pathway TAI trajectories after birth during infancy, childhood, adolescence, and adulthood seems to approximate the general trend of whole-genome-based TAI trajectories that have been reported for other species (***Domazet-Lošo and Tautz, 2010***), more than it resembles the whole-genome-based TAI trajectories in humans. In our gene subset, adolescents expressed more modern OT pathway genes, whereas ancient genes were expressed the strongest in adults. This is a pattern which occurs, too, for instance in zebrafish, Drosophila, and C. elegans. The TAI signature of the oxytocin signaling pathway gene set stands out with a small peak during the prenatal phase, i.e., expression of evolutionary younger genes tends to be slightly increased before birth. Across species, for the entirety of protein-coding genes within a genome, TAIs are usually at a low-point during the prenatal ontogenetic stage. The transcriptome age indices of the OT pathway genes, however, behave somewhat contrary as they culminate during the prenatal period, and then fall. However, the still comparably low mean TAI value of the OT signaling pathway gene set pre-birth (1.66), and the rather small difference of the prenatal mean TAI versus the combined mean TAI across infancy, childhood, adolescence, and adulthood (0.15 units difference), the lack of a statistically significant

difference between ontogenetic stages, with due regard to the overall range/scaling of TAIs in humans, raises the questions of how meaningful this peak at the prenatal stage is. Apart from that, the TAI peak before birth could also be partly accounted for by how the indices are calculated. The raw relative expression values initially give the impression that ancient genes must be predominantly expressed during the prenatal ontogenetic stage over younger genes, and that this pattern appears to reverse for the remaining ontogenetic stages (**supplementary materials 6**). However, since the TAI function systematically weights younger phylostrata stronger than older phylostrata, the TAI signatures we observe likely represent the ages of the oxytocin pathway gene set more realistically. Nevertheless, this is a surprising and interesting finding. Both these and parts of the previous phylostratigraphy analysis results rely on the AHBA brain expression data set, which is unprecedented with regard to resolution and detail of parcellation of brain regions. With this major advantage of precision, however, comes the disadvantage of sample characteristics that the expression data is based on. The AHBA sample consists of six donors only, with mixed ancestries, distinct medical histories, and an unbalanced gender ratio. To address this heterogeneity in the sample, we used the differential stability (DS) measure to quantify the level of consistency and reproducibility of the expression profiles of the initial 154 oxytocin signaling pathway genes across donors and thus of our findings. According to **Hawrylycz and colleagues (2015)**, the 50% threshold of DS scores (i.e., the top 50% DS scores) marks good, reproducible across-donor gene expression patterns. We found that for our gene sample, differential stability scores of 77.21% of the oxytocin pathway genes (105 out of 136, no data available for 18 genes) were among the top 50% of all protein-coding genes, including the three oxytocin candidate genes *OXT*, *OXTR*, and *CD38*. This is indicative of relatively stable expression intensities of the genes of the oxytocin pathway across donors within a comparable age category (suppelementary materials X). Furthermore, for our between- and within-brain-region analyses we decided to parcellate the AHBA data using the Desikan-Killiany (DK) brain atlas (**Desikan et al.,2006**). There are currently almost 20 brain atlas parcellations published, and growing, with widely varying resolutions, most of them focusing exclusively on cortical regions, some of them being designed for sub-cortical structures, the cerebellum, connectivity and network analysis, and more. The availability of such a diverse selection of atlases is a rich resource, but their application may also compromise reproducibility. We chose to implement the DK atlas because it is still among the widest applied (cortical) atlases, thus increasing the reproducibility of our results. In addition, the slightly decreased resolution of the atlas may eventuate in more samples per region, thus yielding a more robust outcome.

We then took a leap in the evolutionary timeline of the oxytocin signaling system and focused on putative patterns of selective pressures in more closely related primate species, from the evolutionary most distant common marmoset over the western gorilla to the modern human. If potential changes in nucleotide sequences of the OT pathway genes in one of these species or genera occurred, that can point to 'recent' events of natural selection throughout the primate lineage and indicate further (functional) adaptions of the gene set within a clade, possibly coinciding with changes in environmental circumstances. The dN/dS ratios of the oxytocin pathway genes, which compared the genome of a primate species to the genome of a common primate ancestor, were, with a few exceptions, consistently below zero, suggesting that the oxytocin pathway gene set has been under evolutionary constraint at least since the emergence of the *anthropoids* from the *haplorhini* approximately 45 - 40 million years ago (**Jäger et al., 2019**), and has presumably undergone only little changes in nucleotide sequences. This would be in line with research suggesting that ancient genes from early phylostrata are under stronger restraint and purifying selection, while more recently emerged genes are less affected by purifying selection and more likely to show adaptive mutations/substitutions (**Cai & Petrov, 2010)**. The clade containing the *Panini* and *Hominini* sister tribes stands out with a slight yet prominent shift of OT pathway gene ω ratios towards neutral selection (i.e., higher dN/dS ratio values; mean ω = 0.23 vs. mean ω = 0.11), which demarcates the four species from the older primate species and could connote minor events of positive selection at or less constraint on some nucleotide sites within the oxytocin pathway at the time *Homininae* fam-

ily split into the *Panini-Hominini* and *Gorillini* tribes (approx. 10 myr; **Langergraber et al., 2012**). Another aspect that demarcates the oxytocin pathway gene dN/dS distribution in the closely related *Pan* and *Homo* genera from the other primates is the presence of some oxytocin pathway genes under clear positive selection. *PPP1R12A* is under beneficial selection in the common chimpanzee. The protein phosphatase 1 regulatory sub-unit 12A emerged in the first phylostratum and accordingly is primarily involved in physiological pathways like muscle contraction, morphogenesis, and cell migration in humans and other species, which is why it has been also coined a developmental gene (**Hughes et al., 2020**). Recently, studies have linked it to different malignant bio-physiological issues, like cancer (**Benuto et al., 2020**; **Zheng et al., 2019**) and hypertension during pregnancy (**Kono et al., 2021**). *MYLK4* and *CALML4* are under positive selection in the Neanderthal, the penultimate archaic human ancestor of the modern human. The two genes originated during the first phylostratum as well. *MYLK4* (myosin light chain kinase family member 4) seems to play a role in growth traits in different animal species (**Shi et al., 2020**; **Zheng et al., 2019**) and, again, in promoting unregulated cell growth in tumors (**Yang et al., 2021**). The functions and implications of *CALML4* (calmodulin like 4) still are uncertain and currently being explored (**Choi et al., 2020**). Prominent is also the drift of *CD38* in the three most diverged primate species, the common marmoset, rhesus macaque, and the sumatran orangutan. The distinctive increase in dN/dS values might be indicative of decreased evolutionary constraint and increased substitutions rates within the boundaries of the gene that is implicated in oxytocin secretion, exclusively in the three primates. However, in common with the divergence of the *Callithrix*, *Macaca*, and *Pongo* clades, the divergence of the *Panini* and *Hominini* clade from the main lineage is an incidence that still is veiled with ambiguities. There are virtually no African fossil remains that could document and reliably give rise to the circumstances surrounding the lineage split of the ancestors of the *Gorilla* clade from the ancestors of the *Pan-Homo* clade (**Feagle, 2013**). Consequently, is unclear whether environmental circumstances were involved in the lineage splits or whether they diverged by genetic drift. Neither is evident which, if any in particular, socio-behavioral or bio-physiological singularities distinguish the ancestors of the *Pan-Homo* clade or the other three primate clades from a common primate ancestor, that could explain the shift of dN/dS ratios of the three oxytocin pathway genes and *CD38*, respectively. A shift due to minor positive selection is conceivable, but due to these limitations must remain at the level of speculation.

Altogether, our findings confirm preceding studies having reported that vasopressin and oxytocin homologues date back at least 550 - 700 million years (**Donaldson & Young, 2008**; **Knobloch & Grinevich, 2014**). Our phylostratigraphic analysis expands those findings, suggesting that the majority of founders of genes that support oxytocin signaling pathway today may date back even further, as far as to the emergence of unicellular life 3.8 - 1.3 billion years ago (**Cooper, 2000**; **Embley & Martin, 2006**). Ancient oxytocin pathway genes likely emerged together with basic cellular organisms and are among other things implicated in muscle contraction, cardio-vascular functions, and metabolism, while the majority of more modern oxytocin pathway genes mainly emerged with the origin of nervous system centralization and is enriched in brain tissue. The substitution rate ratios in the extant eight primate descendants of anthropoids suggest that the ancient oxytocin gene set has largely remained conserved throughout recent primate evolution, while we see slight overall increaes in dN/dS ratios and some genes under positive selection that are primarily implicated in unregulated cell growth and development in the *Pan-Homo* lineage. *CD38* seems to have been under less constraint in more diverged primate species. The transcriptome age signature of the oxytocin signaling pathway genes is comparably low and, in contrast to whole-genome TAI signatures of the human and other species, almost constant across phylogeny. Minor fluctuations tend to resemble the transcriptome age signatures of those species whose ancestors emerged earlier in evolution, rather than the transcriptome age signatures of the human protein-coding genome in the ontogenetic stages after birth, which may be additionally indicative of the gene set's old age, strong conservation, and relevance across species. In humans, more mordern oxytocin genes seem to be slightly more expressed before birth. Today's OT signaling system is implicated in vari-

ous functions in a wide variety of species, ranging from basic physiological functions like gut motility in annelida to a higher-order functions like modulating social interactions in birds and mammals (**Feldman et al., 2016**). Considering the oxytocin signaling pathway gene emergence timeline, the evolutionary development milestones linked to the different phylostrata, and the functional annotation of the ancient and modern OT pathway genes, it seems like the physiological, non-social functions emerged first, adequate to the environmental demands at the time in evolution, and evolved along the constantly changing demands into more sophisticated biological systems. With the major advancement of the emergence of CNS like neuronal organizations during the sixth and seventh phylostratum, we assume that the OT system adapted its functionality again by 'producing' new genes (e.g., *OXTR*, *CD38*, the CACNG-family).

This systematic evolutionary evaluation of the oxytocin signaling pathway genes underpins what has previously been suggested (**Feldman et al., 2016**) and may provide additional evidence for the multi-functionality of the current OT system and it's multi-faceted role in various domains (social, biological, physiological) in mammals, nematoda, amphibians, arthropoda, and more, that we can observe at the present day. Further, the evolutionary, potentially adaptive history and current diverse functionality in a multitude of processes support the in popularity gaining theory that oxytocin could act as a neuromodulator involved in homeostasis and possibly even allostasis (**Carter, 2018**; **Leng & Russel, 2018**; **Quintana & Guastella, 2020**). Additional and especially updated research on the whole oxytocin pathway system in more and diverged species (e.g., cnidarians, porifera, annelida), that compliments the already abundant research in mammals, would be necessary to further substantiate this proposed role of the oxytocin system.

Methods and Materials

The evolutionary signature of the oxytocin pathway

If not stated otherwise, the statistical software R (version 4.0.5; **R Core Team, 2021**) and RStudio (version 1.4.1106; **RStudio Team, 2021**) were used for analyses and data visualizations.

Phylostratigraphy and myTAI

A phylostratigraphic map from Domazet-Lošo and Tautz (***Domazet-Lošo and Tautz, 2008***) with gene age inferences was combined with human gene expression data extracted from the BrainSpan atlas (AHBA; n = 6, 1 female and 5 males, ages 24.0 - 57.0, three of Caucasian ethnicity, two African American, one Hispanic; http://brainspan.org) to compile phylostratigraphy-expression matrices with twelve phylogenetic stages. Data was collected for the complete set of oxytocin signaling pathway genes (n = 154), which was identified using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (http://www.kegg.jp). The full gene list is available in **supplementary materials 1**. In the matching process of the phylostratigaphy map with the expression data, 16 genes were not available, yielding a final sample size of 138 genes. For the descriptive phylostratigraphy, for which the phylostratum and gene count information was of primary interest, two reduced, simplified versions of the phylostratigraphy-expression matrix were extracted, one containing data for the oxytocin genes only, and another matrix with data on all protein-coding genes. RNA sequencing data from the adult brain only was gathered for the left hypothalamus to assess the oxytocin signaling pathway gene count per phylostratum. For subsequent evolutionary trancriptomics using the 'myTAI' R package (version 0.9.3; https://github.com/drostlab/myTAI; (***Drost et al., 2018***)), RNA sequencing data from five ontogenetic stages (prenatal, infant, child, adolescent, adult) and from sixteen brain regions of the BrainSpan atlas (primary motor cortex (M1C), dorsolateral prefrontal cortex (DFC), ventrolateral prefrontal cortex (VFC), orbital frontal cortex (OFC), primary somatosensory cortex (S1C), inferior parietal cortex (IPC), primary auditory cortex (A1C), caudal superior temporal cortex (STC), inferolateral temporal cortex (ITC), primary visual cortex (V1C), medial prefrontal cortex (MFC), hippocampus (HIP), striatum (STR), amygdala (AMY), mediodorsal nucleus of the thalamus (MD), cerebellar cortex (CBC)) were collected. Enrichment of phylostrata in the oxytocin gene set of interest was quantified with the EnrichmentTest() function of the myTAI R

package. At the core of the function operates a Fisher's test that tests for statistical significance of enriched or over-/under-represented phylostrata within a set of genes of interest (like the oxytocin signaling pathway gene set) compared against a set of background genes, for which the full phylostratigraphy-expression matrix was utilized here. TAIs for each brain region across the five ontogenetic stages based on the expression data of 138 genes were computed with the myTAI main function for the calculation and visualization of transcriptome age indices, PlotSignature(). The Transcriptome Age Index (TAI) is calculated for each ontogenetic stage on *n* genes included in the analysis. The phylostratum value (i.e., 1 - 12) of each gene *g* out of *n* genes is multiplied by the corresponding mRNA intensity of gene *g*. The resuling *n* products are then added up and divided by the sum of *n* mRNA intensities. The calculation method of the TAI accounts for the 'overpopulation' of older phylostrata with genes and for the fact that younger phylostrata, on average, contain less genes, so that the index is balanced and weights younger phylostrata increasingly stronger (***Domazet-Lošo and Tautz, 2010***). A statistical permutation test for the deviation from a flat line (***Drost et al., 2018***) was performed for each brain region, to test for any differences in the TAI between ontogenetic stages.

FUMA

To functionally annotate the results from the phylostratigraphy and TAI analyses, the data was submitted to the FUMA web application GENE2FUNC in two separate queries for the "ancient" (genes in phylostrata 1-3) and "modern" (genes in phylostrata 4-12) gene set (MHC excluded; all Bonferroni corrected, FDR correction for the gene-set enrichment testing). The tool uses RNA-seq data from GTEx v6 to test for tissue specific enrichment of the gene set and performs a hypergeometric test to assess enrichment of genes in different categories (i.e. GWAS) and pathways (i.e. calcium or *GNRH* signaling pathways) (***Watanabe et al., 2017***).

Modern gene set brain expression

The FUMA tissue specificity results were extended with between- and within-brain region expression specificity analyses for the modern gene set.

For these, first, AHBA expression data (https://human.brain-map.org, **Hawrylycz et al., 2012**) was pre-processed and prepared for the subsequent analyses with the Python toolbox 'abagen' (version 0.1.1; https://github.com/rmarkello/abagen; **Markello et al., 2021**) in Jupyter Lab (REF) based on the processing pipeline suggested by **Arnatkevičiūtė et al., 2019**. The following processing steps were performed for each of the six donors and cortical and sub-cortical in one joint query. Cortical and sub-cortical expression data for 6 brains was prepared using the 83-region (34 cortical and 8 sub-cortical areas lateralized, one sub-cortical structure bilateral) volumetric Desikan-Killiany atlas in MNI space (**Desikan et al., 2006**). Microarray probes were re-annotated using data provided by **Arnatkevičiūtė et al., 2019**; probes not matched to a valid Entrez ID were discarded. Probes were then filtered based on their expression intensity relative to background noise (**Quackenbush, 2002**). Probes with intensity less than the background in less than 50.00% of samples across donors were excluded, yielding 31,569 probes. When multiple probes indexed the expression of the same gene, the probe with the most consistent pattern of regional variation across donors (i.e., differential stability; **Hawrylycz et al., 2015**) was selected. The MNI coordinates of tissue samples were updated to those generated via non-linear registration using the Advanced Normalization Tools (ANTs; https://github.com/chrisfilo/alleninf). Samples were assigned to brain regions in the Desikan-Killiany atlas if their MNI coordinates were within 2 mm of a given parcel. To reduce the potential for misassignment, sample-to-region matching was constrained by hemisphere and gross structural divisions (i.e., cortex, subcortex/brainstem, and cerebellum, such that e.g., a sample in the left cortex could only be assigned to an atlas parcel in the left cortex (**Arnatkevičiūtė et al., 2019**). If a brain region was not assigned a tissue sample based on that procedure, every voxel in the region was mapped to the nearest tissue sample from the donor in order to generate a dense, interpolated expression map. The average of these expression values was taken across all voxels in

the region, weighted by the distance between each voxel and the sample mapped to it, in order to obtain an estimate of the parcellated expression values for the missing region. Inter-subject variation was addressed by normalizing tissue sample expression values across genes using a robust sigmoid function (**Fulcher et al., 2013**). Gene expression values were normalized across tissue samples using an identical procedure. All available tissue samples were used in the normalization process regardless of whether they were assigned to a brain region. Tissue samples not matched to a brain region were discarded after normalization. Samples assigned to the same brain region were averaged separately for each donor. These processing steps resulted in six expression matrices, one for each donor, with 83 rows corresponding to brain regions and 15,633 columns corresponding to genes. Finally, 19 modern OT pathway genes were extracted (*NFATC2* was not available in the processed matrices). Differential stability (DS) values were also obtained with the 'abagen' toolbox. Because the values are not directly accessible from the above described workflow, they were calculated in an adjunct query with the difference that the DS estimates obtained with this procedure are computed as correlations between parcels in the Desikan-Killiany atlas as opposed to the main query where they are computed between the AHBA-defined brain structures. The DS values from the main query and the adjunct query can thus differ minimally, however they generally are highly similar.

Subsequently, between-brain region and within-brain region comparisons of the modern oxytocin pathway gene set expression profile with different population means were calculated with the prepared matrices. Between-region comparisons were implemented with 42 two-sided one-sample *t*-tests (one *t*-test per brain region, only left cortical hemisphere). For each brain region, the expression intensity of the modern oxytocin gene set of that region was compared against the population mean (i.e., the average expression of the modern oxytocin gene set across all available brain regions); *p*-values were adjusted for 42 comparisons with the FDR correction method. The Desikan-Killiany cortical atlas values were visualized using the R package 'ggseg' (version XXX; https://github.com/ggseg/ggseg; **Mowinckel & Vidal-Piñeiro, 2020**). For the within-region comparisons, 42 two-sided one-sample *t*-tests (one *t*-test per brain region, only left cortical hemisphere) were performed. Here, we investigated whether the expression intensity of the modern oxytocin gene set in a given brain region significantly deviated from the average expression intensity of all available protein-coding genes (n = 15,633, OT genes included) in the same brain region (*p*-values were FDR corrected for 42 tests). As a measure of effect size, Cohen's d for one-sample *t*-tests for both between- and within-region analyses were calculated.

NeuroSynth

Lorem ipsum dolor sit amet, consectetur adipisici elit, sed eiusmod tempor incidunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquid ex ea commodi consequat. Quis aute iure reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint obcaecat cupiditat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Patterns of natural selection across primate species

dN/dS ratios for the oxytocin pathway gene set in the modern human (*Homo sapiens*), two archaic humans (*Homo neanderthalensis*, *Homo denisova*) and five primate species (*Pan troglodytes*, *Gorilla gorilla*, *Pongo abelii*, *Macaca mulatta*, *Callithrix jacchus*) were obtained by submitting a list of the oxytocin pathway genes (n = 154) to the EvoGen online tool (https://genevo.pasteur.fr/; **Dumas et al., 2021**). Only 1-to-1 orthologs and exact matches were selected, which lead to the exclusion of 46 genes from the analysis, yielding a final sample of 107 genes. All allele frequencies were included. As in **Dumas et al., 2021**, we choose medium-quality of gene coverage for the main visualization of the data. Plots for high and low quality coverage are available in the **supplementary materials 4 and 5**, respectively. The phylogenetic tree was visualized with the R package 'ggtree' (version XXX; https://github.com/YuLab-SMU/ggtree; **Yu et al., 2016**). The EvoGen tool draws on data provided

by **Dumas and colleagues (2021)**. In their study, the authors aligned the genomes of the eight primate species with a reference genome of their common primate ancestor and calculated the dN/dS ratios for each of these alignments and orthologous protein-coding gene or DNA sequences. The ratio was computed in the PamL (Phylogenetic Analysis by Maximum Likelihood) program package with the "yn00" algorithm (**Yang, 2007**). Briefly, between all taxa and the ancestor, first the number of synonymous and non-synonymous mutations or substitutions was obtained, then the rate of the number of non-synonymous mutations per non-synonymous site (dN) and the rate of the number of synonymous ("silent") substitutions per synonymous site (dS) was calculated. These two parameters constitute the dN/dS (or Ka/Ks) which measures selective pressure and can be indicative of negative, neutral, and positive selection (**Yang & Bielawski, 2000**; for a detailed description of the methods cf. **Dumas et al., 2021**). Besides, it does not seem to be conclusively agreed upon whether the *Pan* kind should be included in the *Hominini* tribe or whether it is it's own tribe, *Panini*, the sister tribe of the *Hominini*, or whether they 'simply' belong to the shared family *Homininae* instead (**Robson & Wood, 2008**). Recent literature appears to use the terminus *Hominini* exclusively for the *Homo*, *Australopithecus*, and *Paranthropus* kind (**Hawks, 2018**), which we implemented for this paper.

Acknowledgments

This research was funded by Research Council of Norway (301767) and the Novo Nordisk Foundation (NNF16OC0019856). Figure 1A and Figure 3A were partly created with Biorender.com. The phylogenetic tree in ***Figure 6*** was generated with the help of the Mega X software (https://www.megasoftware.net/). The methodological report for the AHBA expression data pre-processing steps was generated with the abagen toolbox (https://github.com/rmarkello/abagen).

Data and code availability

AHBA/BrainSpan data is available for download at ... R and Jupyter Notebook/Python scripts used in the analyses with information on specific parameter settings and additional notes are available at http://www.GitHubLink.com.

Author contributions

XX conceived the study. XX analyzed the data, with contributions from XX. All authors contributed to the interpretation of results. XX wrote the first draft of the paper and all authors contributed to the final manuscript.

References

Domazet-Lošo T, Brajković J, Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. Trends in Genetics. 2007 Nov; 23(11):533–539. https://www.cell.com/trends/genetics/abstract/S0168-9525(07)00299-5, doi: 10.1016/j.tig.2007.08.014, publisher: Elsevier.

Domazet-Lošo T, Tautz D. An Ancient Evolutionary Origin of Genes Associated with Human Genetic Diseases. Molecular Biology and Evolution. 2008 Dec; 25(12):2699–2707. doi: 10.1093/molbev/msn214.

Domazet-Lošo T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature. 2010 Dec; 468(7325):815–818. doi: 10.1038/nature09632.

Drost HG, Gabel A, Liu J, Quint M, Grosse I. myTAI: evolutionary transcriptomics with R. Bioinformatics. 2018 May; 34(9):1589–1590. doi: 10.1093/bioinformatics/btx835.

Feldman R, Monakhov M, Pratt M, Ebstein RP. Oxytocin pathway genes: evolutionary ancient system impacting on human affiliation, sociality, and psychopathology. Biological Psychiatry. 2016; 79:174–184. doi: 10.1016/j.biopsych.2015.08.008.

Guastella AJ, Hickie IB. Oxytocin treatment, circuitry and autism: a critical review of the literature placing oxytocin into the autism context. Biological Psychiatry. 2016; 79:234–242. doi: 10.1016/j.biopsych.2015.06.028.

Hill KR, Walker RS, Božičević M, Eder J, Headland T, Hewlett B, Hurtado AM, Marlowe F, Wiessner P, Wood B. Co-residence patterns in hunter-gatherer societies show unique human social structure. Science. 2011; 331(6022):1286–1289. doi: 10.1126/science.1199071.

Hublin JJ, Spoor F, Braun M, Zonneveld F, Condemi S. A late Neanderthal associated with Upper Palaeolithic artefacts. Nature. 1996; 381(6579):224. doi: 10.1038/381224a0.

Insel TR, Shapiro LE. Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. Proceedings of the National Academy of Sciences. 1992; 89:5981–5985. doi: 10.1073/pnas.89.13.5981.

Jurek B, Neumann ID. The oxytocin receptor: from intracellular signaling to behavior. Physiological reviews. 2018; 98(3):1805–1908. doi: 10.1152/physrev.00031.2017.

Kelava I, Rentzsch F, Technau U. Evolution of eumetazoan nervous systems: insights from cnidarians. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015 Dec; 370(1684):20150065. doi: 10.1098/rstb.2015.0065.

King LB, Walum H, Inoue K, Eyrich NW, Young LJ. Variation in the Oxytocin Receptor Gene Predicts Brain Region–Specific Expression and Social Attachment. Biological Psychiatry. 2015; .

Nielsen C. Six major steps in animal evolution: are we derived sponge larvae? Evolution & Development. 2008; 10(2):241–257. doi: 10.1111/j.1525-142X.2008.00231.x.

Nielsen M. Imitation, pretend play, and childhood: Essential elements in the evolution of human culture? Journal of Comparative Psychology. 2012; 126(2):170–181. doi: 10/dppqm3.

Powell A, Shennan S, Thomas MG. Late Pleistocene demography and the appearance of modern human behavior. Science. 2009; 324(5932):1298–1301. doi: 10.1126/science.1170165.

Quintana DS, Rokicki J, van der Meer D, Alnæs D, Kaufmann T, Córdova-Palomera A, Dieset I, Andreassen OA, Westlye LT. Oxytocin pathway gene networks in the human brain. Nature Communications. 2019 Feb; 10(1):668. doi: 10.1038/s41467-019-08503-8.

Rokicki J, Kaufmann T, Glasø de Lange AM, van der Meer D, Bahrami S, Sartorius AM, K Haukvik U, Eiel Steen N, Schwarz E, Stein DJ, Nærland T, Andreassen OA, Westlye LT, Quintana D. Oxytocin receptor expression patterns in the human brain across development. OSF Preprints. 2021 Apr; https://osf.io/j3b5d/.

Theofanopoulou C, Gedman G, Cahill JA, Boeckx C, Jarvis ED. Universal nomenclature for oxytocin–vasotocin ligand and receptor families. Nature. 2021 Apr; 592(7856):747–755. https://www.nature.com/articles/s41586-020-03040-7, doi: 10.1038/s41586-020-03040-7, number: 7856 Publisher: Nature Publishing Group.

Vaccari C, Lolait SJ, Ostrowski NL. Comparative Distribution of Vasopressin V1b and Oxytocin Receptor Messenger Ribonucleic Acids in Brain 1. Endocrinology. 1998; 139:5015–5033. doi: 10.1210/endo.139.12.6382.

Vaidyanathan R, Hammock EA. Oxytocin receptor dynamics in the brain across development and species. Developmental Neurobiology. 2017; 77(2):143–157. doi: 10.1002/dneu.22403.

Vargas-Pinilla P, Paixão-Côrtes VR, Paré P, Tovo-Rodrigues L, Vieira CMdAG, Xavier A, Comas D, Pissinatti A, Sinigaglia M, Rigo MM. Evolutionary pattern in the OXT-OXTR system in primates: coevolution and positive selection footprints. Proceedings of the National Academy of Sciences. 2015; 112(1):88–93. doi: 10.1073/pnas.1419399112.

Warneken F, Tomasello M. Altruistic helping in human infants and young chimpanzees. science. 2006; 311(5765):1301–1303. doi: 10.1126/science.1121448.

Watanabe K, Taskesen E, Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017; 8:1826. doi: 10.1038/s41467-017-01261-5.

Winslow JT, Insel TR. The social deficits of the oxytocin knockout mouse. Neuropeptides. 2002; 36:221–229. doi: 10.1054/npep.2002.0909.

Winterton A, Westlye LT, Steen NE, Andreassen OA, Quintana DS. Improving the precision of intranasal oxytocin research. Nature Human Behaviour. 2020 Nov; p. 1–10. doi: 10.1038/s41562-020-00996-4.