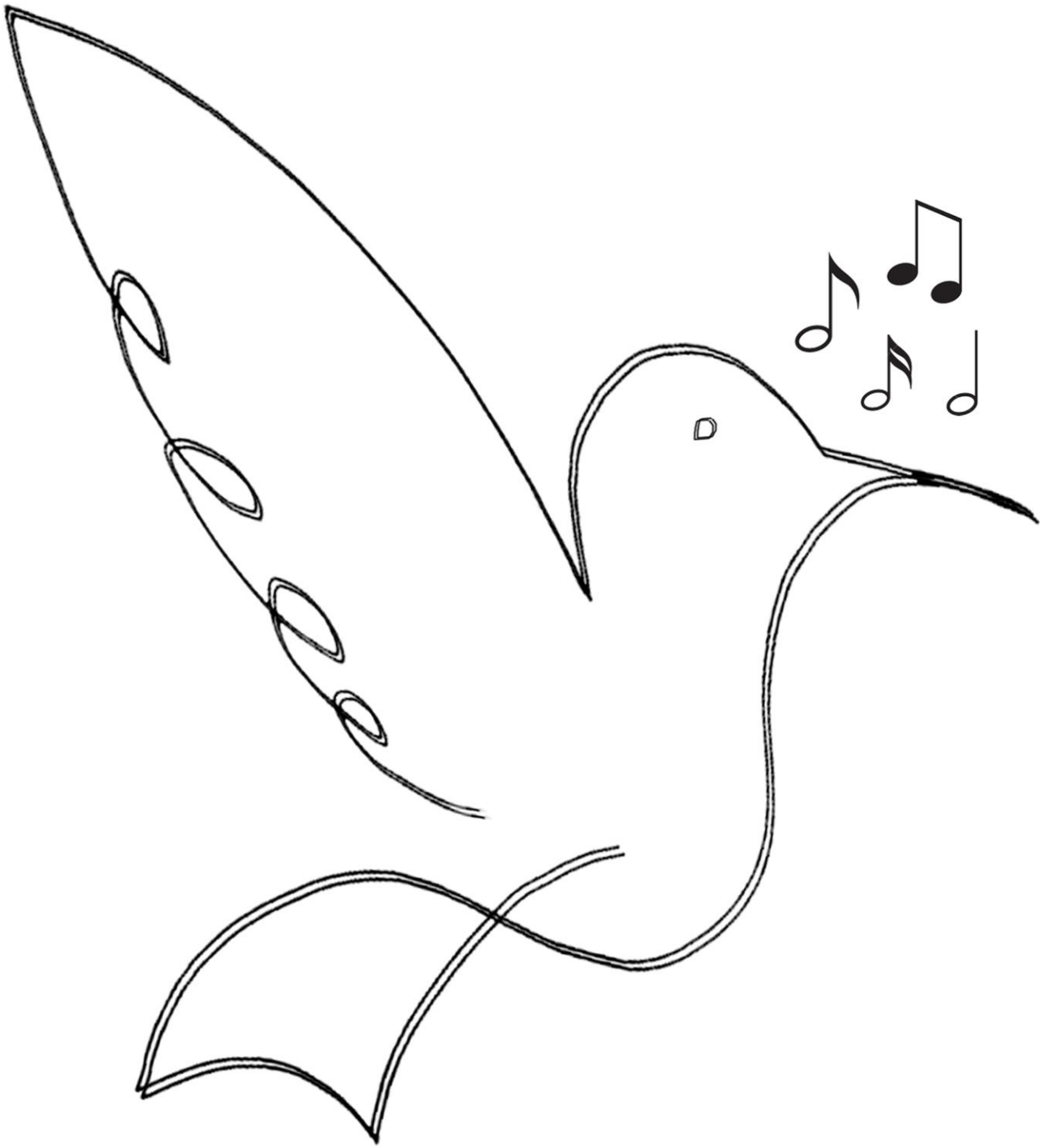
Implications of oxytocin in speech

Constantina Theofanopoulou

Implications of oxytocin in speech

Constantina Theofanopoulou



PhD program: Cognitive Science and Language
University of Barcelona

Advisor-Tutor: Cedric Boeckx (University of Barcelona)
Co-advisor: Erich D. Jarvis (Rockefeller University)

Table of contents

Acknowledgments

To my family and friends
Who'd meet their ends
to make ends meet
All those who shed a tear, shared a beer, helped my mind be in gear
Rode with me in craziness and heaven
and love
Who my writings read, proofread
Got me out of the red
Danced flamenco dances
All in situ
in synchrony and synteny
Gave me a hand
to make ends meet
they'd meet their ends
my family and friends.

Abstract

The overall aim of my thesis is to shed light on the interplay between the evolution of human sociality and the evolution of human language. My approach is multidisciplinary and includes studies ranging from genomic analyses in humans and other species to behavioral experiments in songbirds. Findings presented here lend genetic evidence to a specific hypothesis under which human sociality can be studied, the 'self-domestication' hypothesis, based on significant overlaps identified in the genes under positive selection in modern humans and several domesticated species. We further propose oxytocin as a good candidate molecule that underpins the genetic mechanisms of human prosociality and language, by serving as the molecular basis of social reward in spoken language acquisition. We show modern human-specific alleles in the oxytocin and the paralogous vasopressin/vasotocin receptors (particularly *AVPR1A/VTR1A*) correlate with a shift towards prosociality in modern-humans, along with three convergent variant changes in modern humans and bonobos, who have also been claimed to be self-domesticated. Additionally, we show an effect of social reward in tuning fine-grained aspects of vocal learning (i.e. pitch learning) in an experiment in zebra finches, providing support for the hypothesis on the importance of social feedback in human spoken-language acquisition. We also present preliminary findings on the role of oxytocin in singing in zebra finches; male zebra finches treated intranasally with an oxytocin-antagonist reduced significantly the number of introductory notes in the song they sang to attract females. Lastly, we propose a universal nomenclature for the vertebrate oxytocin and vasopressin/vasotocin ligands and receptors, which is based on multi-scale synteny analyses. The nomenclature will allow easier translation of findings across vertebrates and foster more informative design of experiments across species.

Introduction

In my thesis I investigate the evolution of human sociality and its possible role for providing a scaffold to the evolution of language (Kuhl, 2007; Tomasello, 2003). My approach is multidisciplinary and includes studies ranging from genomic analyses to behavioral experiments in songbirds. My hypothesis is that oxytocin is a good candidate molecule that could help us decipher the role the evolution of our sociality plays in the evolution of language, as well as the role of social reward and social motivation in language acquisition. To address this hypothesis, I also focus on the evolution of the oxytocin and vasopressin/vasotocin gene families, to clarify the correct gene-orthologs we should target to study in non-human species, including making my and others' future findings in songbirds readily translatable to humans.

I study the evolution of human sociality mainly through the lens of the 'self-domestication' hypothesis, according to which natural selection in humans favored increased prosociality over aggression (Hare 2017), giving rise to a behavioral phenotype that is reminiscent of the one we witness in domesticated species. Traits that modern humans show when compared to their cousins, the Neanderthals, are also highly reminiscent of traits that have been independently found to be characteristic of domesticated species when compared to their wild counterparts (reduced ears, shorter muzzles, smaller teeth; smaller cranial capacities, paedomorphosis; reduction of sexual dimorphism (feminization)).

In Theofanopoulou et al. 2018 (Chapter 1) we examine if this old hypothesis of human evolution, stemming from thoughts formulated in Darwin 1888, makes sense at a genetic level. We tested the possible overlap between the genes found under positive selection in modern humans and the genes under positive selection in one or more domesticated species (dog, cat, horse and taurine cattle). We identify a statistically significant intersection of these genes, and further that this intersection remains significant when we compared the genes under positive selection in modern humans against only dog or only cattle. The genes that overlap are genes involved in neural crest formation, synaptic plasticity, memory, and learning, and some of them, when disrupted, can lead to a broad range of syndromes comprising craniofacial defects and cognitive deficits. These findings shed light on the possible genetic underpinnings of the enhanced sensory motor and learning abilities encountered in some domesticated species, another hypothesized by-product of the domestication process (Hare, 2017), and possibly to the underpinnings of human complex communicative abilities.

Oxytocin and its receptor (*OXT/OT* and *OXTR/OTR*) are among the genes that have been studied the most in the context of domestication: different gene expression, methylation and selection patterns have been identified in the domesticates in comparison to their wild cousins (Bence et al., 2017; Fam et al., 2018; Ruan & Zhang, 2016). This along with the array of studies on the role of oxytocin in social cognition (Meyer-Lindenberg, Domes, Kirsch, & Heinrichs, 2011) led us to hypothesize that variant changes in the oxytocin receptor (*OTR*) and the arginine-vasopressin receptors (*AVPR1A* and *AVPR1B*; also called here *VTR1A* and *VTR1B*) between modern humans and our extinct (Neanderthals and Denisovans) and extant relatives (macaques, bonobos, chimpanzees) could have been responsible for our human prosociality (Theofanopoulou, C., Andirko, A., & Boeckx 2018, Appendix Chapter 1). We report 29 variants which were clustered based on their presence in the species studied (e.g. modern human-specific, Homo-specific); all of these variants are sites of Single Nucleotide Polymorphisms in modern humans, hence we were able to retrieve information also on the functional importance of these sites. Regarding modern human-specific alleles, we found one allele in *AVPR1A* (rs11174811) at high frequency and linked to prosocial phenotypes in modern humans, while the ancestral allele is associated with antisocial phenotypes. We also report three sites in *AVPR1A* of putatively convergent changes between modern humans and bonobos

(rs237897(A), rs2228485(G) and rs1042615(A)), not present in chimpanzees. We consider this last finding highly pertinent to the self-domestication process that also bonobos, apart from humans, have been claimed to have gone through in their evolution (Hare, Wobber, & Wrangham, 2012).

In Theofanopoulou 2016 (Chapter 2) I put together studies from the literature that point to a role of oxytocin in modulating the multimodality that characterizes our linguistic ability. I follow a bottom up approach, starting off from possible genetic interactions that could support this role, for example a hypothesized interaction of *OT* and *FOXP2* through *CNTNAP2* or *LNPEP*, and ending with evidence from EEG (electroencephalography) and behavioral studies in humans showing the effect of oxytocin in online visuomotor processing. These studies support oxytocin's proposed role in social interactions of turn-taking during online language production and processing.

In Theofanopoulou, Boeckx, and Jarvis 2017 (Chapter 3), we further present a specific hypothesize for a role of oxytocin in the social motivation for vocal learning, a specialized component of language, that is also find in other species, like songbirds. We build upon molecular findings on the expression patterns of *OT* and *OTR* in the brains of vocal learners and non-(or rudimentary)-vocal learners. We suggest that *OT* and the circuits it functions in are good candidates for the long-hypothesized motivation and reward mechanism of vocal learning. We propose specific neural mechanisms through which *OT* could modulate brain regions that are specialized for vocal learning directly, or indirectly through its interaction with dopaminergic neurons. We lastly propose specific experiments through which such a hypothesis could be tested in songbirds.

In Theofanopoulou et al. (Appendix Chapter 2) we experimentally address the traditional idea that social reward enhances learning in the realm of vocal learning. Even though there are studies showing the strong effect of social feedback in language acquisition (Kuhl, 2007), in human studies it is not possible to dissociate social reward from vocal learning and thus to study the exact impact of social reward on vocal learning. We attempted such a dissociation, developing a vocal learning behavioral paradigm with and without social reward, and tested it in zebra finches, a vocal learning songbird. Juvenile male zebra finches were first operantly taught to imitate a two-syllable song. Then they were exposed to two different contexts, switched every other day: a social reward context, in which an animal male model of a bird and a non-singing but live female bird were present; and a social isolation context with no model or live birds present. In both contexts, the juveniles were exposed to operantly elicited playbacks of one of two very similar songs, comprised of two syllables, the same syllables of the song they had learnt, but differing by two semitones in the pitch of the second syllable. Five out of the six birds tested imitated the pitch of the song they heard in the social reward context, suggesting that finely tuned aspects of vocal learning, like pitch, can be gated by social reward.

In our next experiment (Theofanopoulou et al., Appendix Chapter 3), we sought to test the hypothesis we proposed in Theofanopoulou, Boeckx, and Jarvis 2017, by manipulating the oxytocin-system in zebra finches and studying the effects in their singing. We administered an oxytocin antagonist intranasally in zebra finch males, and then co-housed them with a female to elicit singing to her (directed-singing). We show that oxytocin-antagonist-treated males had a significant drop in the number of introductory notes in their directed love song, more similar to the levels found in undirected song without a female. We also demonstrate that intranasal administration of an oxytocin-antagonist crosses the blood-brain barrier in zebra finches.

Our last study (Theofanopoulou et al. Appendix Chapter 4) came about as a result of our attempt to first, single out which receptor in the avian genomes corresponded to the mammalian OTR, and second, to search for evidence that the system we manipulated in the aforementioned

experiment in songbirds is indeed the oxytocin system, and not a different one (the 'mesotocin' system). We soon realized that oxytocin and vasopressin/vasotocin ligands and receptors appear in the literature with as many names as the classes of vertebrate and invertebrate species whose genome has been sequenced, something that hinders the translation of findings across species. Our goal ended up being to clear up this confusing nomenclature by studying in detail the synteny (genomic territory) of these ligands and receptors in 33 vertebrate genomes that span all major vertebrate lineages and 4 invertebrate outgroups. Our findings indicate that oxytocin and vasopressin/vasotocin are adjacent paralogous genes that formed as a local genomic duplication event near the origin of vertebrates. What has been called mesotocin, isotocin, or oxytocin-like in non-mammalian species are all the same gene, namely oxytocin; vasotocin in all non-mammalian vertebrates is the same as vasopressin in mammals. Thus, following the standard practice in molecular biology, we propose that these two genes be given the same orthologous names across vertebrates and paralogous names relative to each other, namely oxytocin and vasotocin. Additionally, through multi-scale synteny analyses, we clarified the orthology and paralogy of all oxytocin and vasotocin receptors in all major vertebrate classes and we propose a new universal vertebrate nomenclature for them too. We traced their evolutionary history and propose that these receptors formed through a series of duplications: first after one round of whole-genome duplication in the common ancestor of all vertebrates; followed by a segmental duplication in the common ancestor with cyclostomes/lampreys; and then two further segmental duplications in the gnathostome-ancestor and in the osteicthyan-ancestor.

In conclusion, considering all the above, I believe that this thesis offers a fertile ground for future experiments seeking to unravel the effect of social reward in vocal learning, something that can shed light to the effect that evolution of our sociality might have had in the evolution of a fully-fledged language in our species. My thesis also lends evidence to a specific hypothesis under which our sociality can be studied, the 'self-domestication' hypothesis, which in turn opens up a venue for testing possible convergences between humans and domesticates at several levels of biological analyses. Further, the oxytocin and vasotocin systems are shown to be good candidates for uncovering changes that might have had an effect on the evolution of prosociality (cf. convergent variant-changes in modern humans and bonobos), but also changes that affect vocal learning behaviors (cf. differences in singing after oxytocin antagonist administration). Lastly, my thesis proposes a universal nomenclature for the vertebrate oxytocin and vasotocin ligands and receptors, which will allow easier translation of findings across vertebrates and foster more informative design of functional experiments across species.

References

Bence, M., Marx, P., Szantai, E., Kubinyi, E., Ronai, Z., & Banlaki, Z. (2017). Lessons from the canine Oxtr gene: populations, variants and functional aspects. *Genes, Brain and Behavior*, *16*(4), 427–438. https://doi.org/10.1111/gbb.12356

Darwin, C. (1888). *The descent of man and selection in relation to sex*. Murray.

Fam, B. S. O., Paré, P., Felkl, A. B., Vargas-Pinilla, P., Paixão-Côrtes, V. R., Viscardi, L. H., & Bortolini, M. C. (2018). Oxytocin and arginine vasopressin systems in the domestication process. *Genetics and Molecular Biology*, *41*(1 suppl 1), 235–242. https://doi.org/10.1590/1678-4685-gmb-2017-0069

Hare, B. (2017). Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annual Review of Psychology*, *68*(1), 155–186. https://doi.org/10.1146/annurev-psych-010416-044201

Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour*, *83*(3), 573–585. https://doi.org/10.1016/j.anbehav.2011.12.007

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*(1), 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

Meyer-Lindenberg, A., Domes, G., Kirsch, P., & Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine. *Nature Reviews Neuroscience*, *12*(9), 524–538. https://doi.org/10.1038/nrn3044

Ruan, C., & Zhang, Z. (2016). Laboratory domestication changed the expression patterns of oxytocin and vasopressin in brains of rats and mice. *Anatomical Science International*, *91*(4), 358–370. https://doi.org/10.1007/s12565-015-0311-0

Theofanopoulou, C., Andirko, A., & Boeckx, C. (2018). Oxytocin and Vasopressin Receptor variants as a window onto the evolution of human prosociality. *BioRxiv*, *460584*. doi: http://dx.doi.org/10.1101/460584. (Appendix Chapter1)

Theofanopoulou, C. (2016). Implications of oxytocin in human linguistic cognition: From genome to phenome. *Frontiers in Neuroscience*, *10*(271). https://doi.org/10.3389/fnins.2016.00271 (Chapter 2)

Theofanopoulou, C., Boeckx, C., & Jarvis, E. D. (2017). A hypothesis on a role of oxytocin in the social mechanisms of speech and vocal learning. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1861), 20170988. https://doi.org/10.1098/rspb.2017.0988 (Chapter 3)

Theofanopoulou, C., Gastaldon, S., O'Rourke, T., Samuels, B. D., Tiago Martins, P., Delogu, F., … Boeckx, C. (2017). Self-domestication in homo sapiens: Insights from comparative genomics. *PLoS ONE 12*(10): e0185306. https://doi.org/10.1371/journal.pone.0185306 (Chapter 1)

Theofanopoulou, C., Lipkind, D., Tchernichovski, O., Boeckx, C., & Jarvis, E.D. (Appendix Chapter 2). Selective vocal learning in a social reward context.

Theofanopoulou, C., Boeckx, C., & Jarvis, E. D. (Appendix Chapter 3). Pilot study: testing the effect of intranasal administration of an oxytocin-receptor antagonist in adult zebra finch directed singing.

Theofanopoulou, C., Gedman, G., Cahill, J. A., Boeckx, C., & Jarvis, E.D. (Appendix Chapter 4). A proposed universal nomenclature for the oxytocin and vasotocin ligand and receptor families and their evolutionary history.

Tomasello, M. (2003). *Constructing a language: A usage-based approach to child language acquisition.* MA: Cambridge.

Published Chapters

Chapter 1

PLOS ONE

RESEARCH ARTICLE

Self-domestication in *Homo sapiens*: Insights from comparative genomics

Constantina Theofanopoulou[1,2☯], Simone Gastaldon[1,3☯], Thomas O'Rourke[1], Bridget D. Samuels[4], Angela Messner[1], Pedro Tiago Martins[1], Francesco Delogu[5], Saleh Alamri[1], Cedric Boeckx[1,2,6]*

1 Section of General Linguistics, Universitat de Barcelona, Barcelona, Spain, **2** Universitat de Barcelona Institute for Complex Systems, Barcelona, Spain, **3** School of Psychology, University of Padova, Padova, Italy, **4** Center for Craniofacial Molecular Biology, University of Southern California, Los Angeles, CA, United States of America, **5** Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway, **6** ICREA, Barcelona, Spain

☯ These authors contributed equally to this work.
* cedric.boeckx@ub.edu

Citation: Theofanopoulou C, Gastaldon S, O'Rourke T, Samuels BD, Messner A, Martins PT, et al. (2017) Self-domestication in *Homo sapiens*: Insights from comparative genomics. PLoS ONE 12(10): e0185306. https://doi.org/10.1371/journal.pone.0185306

Editor: Michael Klymkowsky, University of Colorado Boulder, UNITED STATES

Received: May 24, 2017

Accepted: September 11, 2017

Published: October 18, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: CB acknowledges the financial support from the Spanish Ministry of Economy and Competitiveness (grant FFI2016-78034-C2-1-P), a Marie Curie International Reintegration Grant from the European Union (PIRG-GA-2009-256413), research funds from the Fundació Bosch i Gimpera, and from the Generalitat de Catalunya (2014-SGR-200). CTh and TOR acknowledge

Abstract

This study identifies and analyzes statistically significant overlaps between selective sweep screens in anatomically modern humans and several domesticated species. The results obtained suggest that (paleo-)genomic data can be exploited to complement the fossil record and support the idea of self-domestication in *Homo sapiens*, a process that likely intensified as our species populated its niche. Our analysis lends support to attempts to capture the "domestication syndrome" in terms of alterations to certain signaling pathways and cell lineages, such as the neural crest.

Introduction

Recent advances in genomics, coupled with an ever-richer body of palaeoarchaeological, anatomical, and animal behavior literature, offer new opportunities to test long-standing hypotheses about human evolution. In the domain of human cognition, the retrieval of ancient DNA can, with the help of well-articulated linking hypotheses connecting genes, brain, and cognition, shed light on the emergence of 'cognitive modernity'. It is to this end that we present data from (paleo-)genomics in support of an old hypothesis about the evolution of our species: that of self-domestication. As has been well documented elsewhere [1, 2], the idea that anatomically modern humans (AMH) are a domesticated species has long been entertained by preeminent scholars in biological and human sciences (in passing by Charles Darwin [3] and more seriously by Franz Boas [4]). We argue that such characterizations are accurate, not merely as analogies, but in identifying shared evolutionary trajectories, with accompanying convergent signatures of selection, in AMH and domesticated species.

In order to explore whether our species is self-domesticated, we must first address what it means to be domesticated and whether AMH meet these criteria. We take the view, defended in more detail elsewhere [1, 5–7], that domesticated species are best categorized in terms of the phenotypic traits that they broadly share, rather than in terms of human mastery, design, or

support from the Generalitat de Catalunya in the form of doctoral (FI) fellowships.

Competing interests: The authors have declared that no competing interests exist.

orchestration. There are inherent weaknesses in the human-mastery or conditions-based views of domestication that an account based on phenotypic traits does not face. The commonly shared traits of domesticates provide the strongest and most objective means by which these animals can be considered a single category. Furthermore, there is now evidence that many of the phenotypic traits of domesticates emerge independently of any human predispositions, intentional or otherwise [7, 8]. A broad consensus is now emerging that "commensal" and "mutualistic" processes can lead to domestication [6, 9–11], whereby both the domesticator and domesticated species seek out and benefit from cohabitation; thus, AMH were not the sole agents in all domestication events. Many of the species that have ultimately come to inhabit domestic niches are widely considered to have done so largely autonomously; in other words, to have self-domesticated. Changes in their social ecology (i.e., both their feeding niche and social organization), along with other parameters, have been recently suggested to confirm this hypothesis [12]. It has been proposed that dogs, cats, foxes [5, 7, 11, 13, 14], and even livestock species such as pigs, sheep, and cattle [6, 11, 15], may have undergone such processes.

Domesticated species display a range of anatomical and behavioral phenotypes that set them apart from their wild counterparts: depigmentation; floppy, reduced ears; shorter muzzles; curly tails; smaller teeth; smaller cranial capacities (and concomitant brain size reduction); paedomorphosis; neotenous (juvenile) behavior; reduction of sexual dimorphism (feminization); docility; and more frequent estrous cycles. Of course, not all of these characteristics are found in all domesticates, but many of them are indeed present to some extent in each [16]. This constellation of features has been referred to as the "domestication syndrome" and has been hypothesized to arise from a mild deficit of neural crest cells [17]. A critical question for the present study is whether our species displays some or all of the phenotypes associated with the domestication syndrome, thus warranting comparison to determine signatures of selection shared with domesticates. Such signatures of domestication can be detected through comparisons of a domesticated species with "either their direct wild-living ancestor or close relatives if the ancestor is no longer extant" [2]. In the case of AMH, since there is no wild extant counterpart available, the obvious comparanda include our closest living relatives (i.e., the great apes) and extinct species of the genus *Homo*, to the extent that relevant data can be extracted from the fossil record.

Many of the anatomical changes associated with domestication describe some of the well-known anatomical differences between AMH and Neanderthals (see Fig 1). The two species display different ontogenetic trajectories [18, 19] resulting in craniofacial differences that invariably lead to a more 'gracile', 'juvenile' profile in AMH relative to Neanderthals. It is well-established that prognathism is significantly reduced in our species [19, 20]. Brow ridges and nasal projections are smaller in AMH than in our most closely related (extinct) relatives [21], as are our teeth [22, 23] and our cranial capacity [24]. This profile is sometimes called 'feminized' [21], and is associated with an overall reduction of sexual dimorphism, which is also associated with domestication [25]. The process of 'feminization' (reduction of androgen levels and rise in estrogen levels [21]) is often associated with reduced reactivity of the hypothalamus-pituitary-adrenal axis [26], a physiological trait thought to be critical for domestication [17, 27]. Evidence from digit ratio comparisons—a measure of prenatal androgen exposure [28]—further suggests that Neanderthals had higher prenatal androgen exposure than AMH [29]. Additional differences in other traits associated with domestication may exist, but there are either obvious confounding factors involved (e.g., geography for pigmentation), or the data are more controversial (as in the case of reproductive cycle changes [30]).

In light of these differences, we contend (*contra* [21]) that self-domestication coincided with the emergence of AMH (*sensu* [31]: specimens sharing a significant number of derived

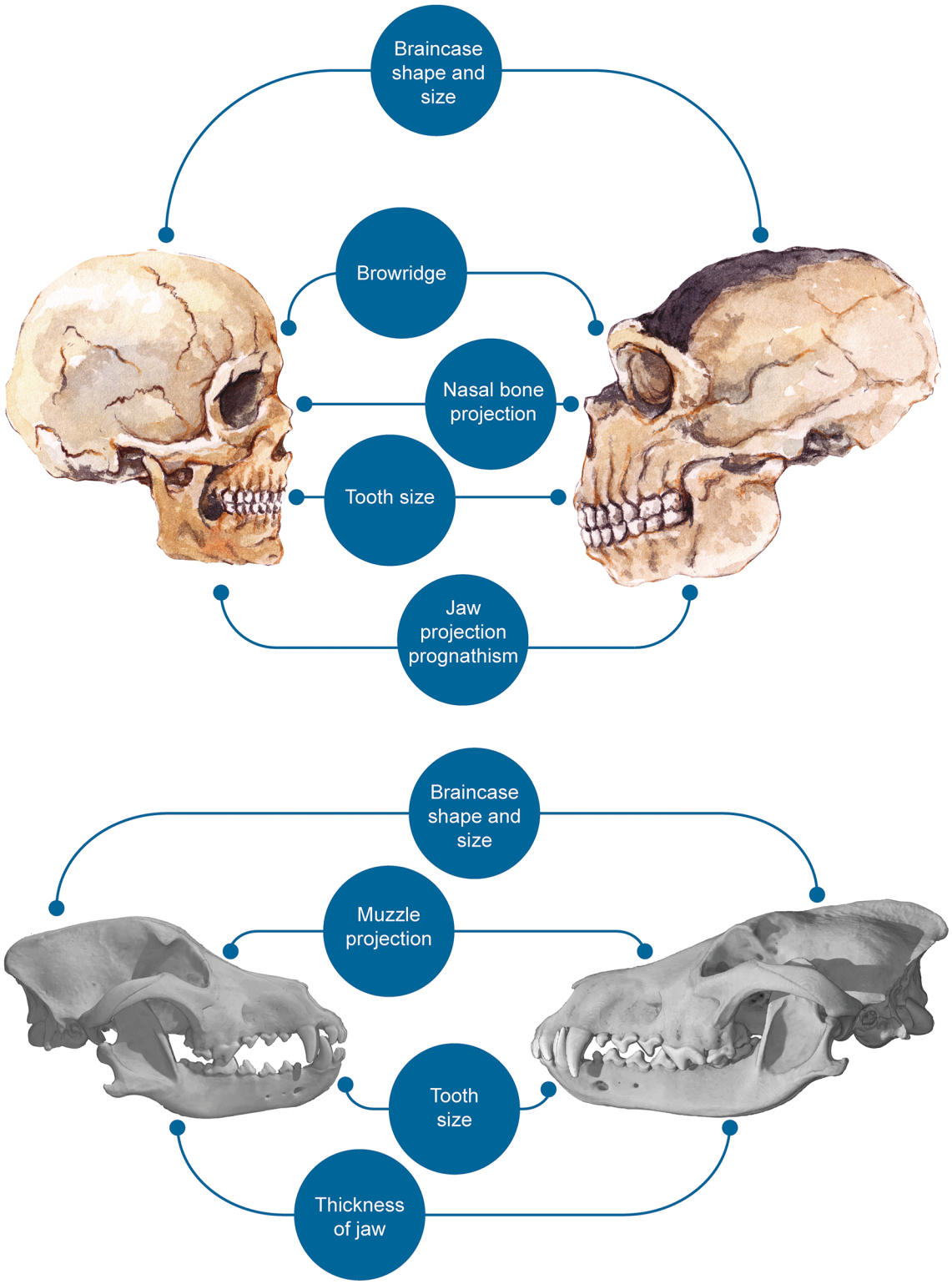


Fig 1. Salient craniofacial differences between AMH and Neanderthals (top) and between dogs and wolves (bottom).

https://doi.org/10.1371/journal.pone.0185306.g001

features in the skeleton with extant members of our species), since the critical phenotypic changes are already present in the first specimens, although this self-domestication process may have intensified as our species expanded geographically and demographically.

Having laid out the case that AMH exhibit characteristics typical of the domestication syndrome, it remains to be clarified how our self-domestication event may have occurred. An obvious difference between AMH and other (self-)domesticated species is that the selective pressures leading to our domestication must have been intraspecific, although it has been suggested that the bonobo (*Pan paniscus*), a species that displays some of the traits of the domestication syndrome, has undergone a similar self-domestication process [25]. But even *inter*specific domestication events suggest that the selective pressures for our self-domestication need not have been qualitatively different from those experienced by other species. The recent domestication of the silver fox (*Vulpes vulpes*) demonstrates this: In the experimental breeding program started by Dmitry Belyaev [7, 8, 26, 32], foxes were intensively selected and bred over more than half a century based on only one criterion, tameness towards humans. Within twenty years of selection for this trait, a range of traits typical of the domestication syndrome had emerged [8]. Crucially, this suggests that selection for tameness is enough to bring about a constellation of domestic traits (see [33]), many of which humans share. The domesticated traits exhibited by AMH plausibly emerged following similar intraspecific selective pressures for prosocial behaviors: in other words, tameness towards fellow humans. Similarly, it has been claimed that reduced emotional reactivity and increased prosociality among humans were keys to our self-domestication [34]. So, what, if anything, differentiates prosociality from self-domestication? Certainly, reduced reactivity or increased prosocial behaviors seem to be necessary precursors of self-domestication, but these are not sufficient to describe the full-blown suite of traits associated with the domestication syndrome. Only consistent selection for such behaviors has been shown experimentally to bring about the far more extensive phenotype of domestication (i.e., in the silver fox experiment), although selection for tameness exclusively does not seem to be the only pressure at work in some cases of domestication (cf. the 'socioecological' factor that may have shaped dog domestication [12]).

Intriguingly, there is evidence that domestication can enable the development of complex behaviors beyond those discussed so far for the domestication syndrome. For example, both dogs and domesticated foxes outperform all non-human primates in tests of cooperative communication [34]. The Bengalese finch, domesticated from its wild ancestor, the white-rumped munia [35, 36], has developed a complex song that is preferred by both female finches and munias over the stereotyped song of the male munia [37]. There are tempting parallels to be drawn here regarding the potential effects of self-domestication on the emergence of human language, relating to the emergence of a fully modern 'language-ready' brain [38–40], or the triggering of our capacity for complex iterative learning, necessary for the cultural transmission of language [2, 41].

The self-domestication hypothesis is, then, a strong contender to account for key aspects of modern human cognition. The central claim of the present paper is that (paleo-)genomic data can provide evidence to complement the anatomical and behavioral data outlined above, which suggest that AMH underwent a process of self-domestication. Crucially, we now have high-quality genomes for our closest extinct relatives, the Neanderthals and Denisovans, allowing for genomic comparison with AMH [42], as well as genomes of several domesticated species, which can be compared with their wild counterparts [43]. This information offers the opportunity to test for the existence of significant overlapping regions showing signatures of positive selection and putatively associated with (self-)domestication.

Results

We examined the overlap of gene sets independently claimed to be under positive selection in AMH (when compared with Neanderthal/Denisovan) and several domesticates for which detailed genetic information is available: dog (*Canis familiaris*), cat (*Felis catus*), horse (*Equus caballus*) and taurine cattle (*Bos taurus*). The pool of domesticates chosen yielded a total of 691 genes, and the total AMH pool, 742 genes. The intersection of these lists was found to be the 41 genes shown in Table 1, which represent all of the genes associated with loci under positive selection both in AMH and in one or more domesticates. A hypergeometric intersection test revealed that the intersection size of 41 was statistically significant ($p < 0.01$). The results are represented graphically in Fig 2; for further details, see S1 and S2 Tables.

We confirmed the significance of this result with a Monte Carlo simulation of 1,000,000 trials, in which samples of 691 and 742 genes were randomly selected with no replacement from a pool of 19,500 (the approximate average number of genes in the genomes of these species). The simulation confirmed that an intersection size greater than or equal to 41 is highly significant ($p = 0.0033$).

To validate that the genes with evidence for positive selection in multiple species are orthologous (rather than simply paralogous) across the species studied, we performed synteny analysis for all 41 genes with evidence of selective sweeps in both AMH and at least one domesticate (Fig 3). We found that all 41 genes are located in syntenic blocks across the species studied (see S6 Table), meaning that the intersection size identified does not include any false positives (i.e., paralogous genes that have been given the same name due to high sequence identity). The same holds for genes associated with loci under selection in multiple domesticated species but not AMH; see again S6 Table.

This situation contrasted with the modest (statistically insignificant) overlaps between the domesticates and several Great Apes for which selective sweep screens were available: chimpanzee (*Pan t. troglodytes*), orangutan (*Pongo abelii*), and gorilla (*G. g. gorilla*) (see S4 Table).

Intersections between domesticates (15 genes in total, see S1 Table) were tested, with a hypergeometric intersection test showing a significant overlap between genes under selection in the dog and in cattle ($p < 0.01$). Furthermore, tests between AMH and each domesticate showed significant overlaps with the dog ($v = 15$, $p < 0.05$) and with cattle ($v = 9$, $p < 0.01$). In order to investigate whether these significant overlaps provide evidence for a convergent effect of domestication, we compared the pool of genes putatively under selection in AMH with genes reported to be under selection in the Eurasian wolf (*Canis lupus lupus*) and wisent (or European bison, *Bison bonasus*). These are the closest related non-domesticated species to the dog and cattle for which there are published studies of genes under selection in modern populations [53–61]. Neither wisent nor wolf populations showed any significant convergence of genes under selection with AMH (see S5 Table). This control comparison suggests that the significant overlap between AMH and both dog and cattle may be an effect of convergent domestication processes in these species.

Since we pooled data concerning positive selection and selective sweeps in AMH from different sources, intersection tests were carried out between the domestication pool and the pool of each AMH dataset used in this study. A significant intersection was found with the data from Prüfer et al. [44] ($p < 0.05$) and with the combined data from Prüfer et al. [44] and Racimo [48] ($p < 0.05$).

Though no gene was found to be shared across all domesticated species studied here as well as AMH, this is not necessarily expected. As discussed in the Introduction, domestication is known to proceed through various routes, and is thus not a uniform affair. However, common pathways can be identified, as can genes that may have contributed to domestication events

Table 1. List of 41 overlapping genes with evidence of positive selection in AMH and domesticated species (for more details, see S2 Table).

Gene name	Overlapping species	Sources of AMH data	Sources of domesticate data
AMBRA1	horse	[44]	[45]
BRAF	cat, horse	[46]	[45, 47]
CACNA1D	horse	[48]	[45]
COA5	dog	[48]	[49]
COL11A1	dog	[46]	[50]
COQ10B	dog	[44]	[50]
DLGAP1	horse	[46]	[45]
ERBB4	cattle	[46]	[51]
FAM172A	cattle, dog	[48]	[50, 51]
GGT7	dog	[46]	[49]
GRIA1	cat	[46]	[47]
GRIK3	dog, cattle	[46]	[50]
HSD3B7	cat	[46]	[47]
HSPD1	dog	[44]	[50]
HSPE1	dog	[44]	[50]
ITGA9	cat	[48]	[47]
LRP1B	cattle	[46]	[51]
LYST	dog	[46]	[49]
MOB4	dog	[44]	[50]
MYLK3	cat	[46]	[47]
NCOA6	dog	[46]	[49]
NEK4	cat	[48]	[47]
NT5DC2	horse	[48]	[45]
NTM	horse	[46]	[45]
PLAC8L1	cat, cattle	[46]	[47, 51]
PPAP2A	cat	[48]	[47]
PPAPDC1B	cat	[44]	[47]
PRR11	cat	[48]	[47]
PVRL3	cattle	[48]	[51]
RFTN2	dog	[44]	[50]
RNPC3	cat, dog	[46]	[47, 50, 52]
SF3B1	dog	[44]	[50]
SKA2	dog	[48]	[49]
SNRPD1	cattle	[44, 46, 48]	[51]
STAB1	horse	[48]	[45]
SYTL1	cat	[48]	[47]
TAS2R16	cattle	[46]	[51]
TEX14	cat	[48]	[47]
TP53BP1	cat	[48]	[47]
ZMYND10	cat	[48]	[47]
ZNF521	cattle	[46]	[51]

https://doi.org/10.1371/journal.pone.0185306.t001

that we think deserve special attention. Five genes were found to be associated with signals of positive selection in AMH and multiple domesticated species (see S3 Table): *RNPC3, FAM172A, PLAC8L1, GRIK3* and *BRAF*.

RNPC3 shows evidence of positive selection in the dog, cat, and AMH. *RNPC3* is one of only two genes with more than one putatively causal variant fixed between dogs and wolves

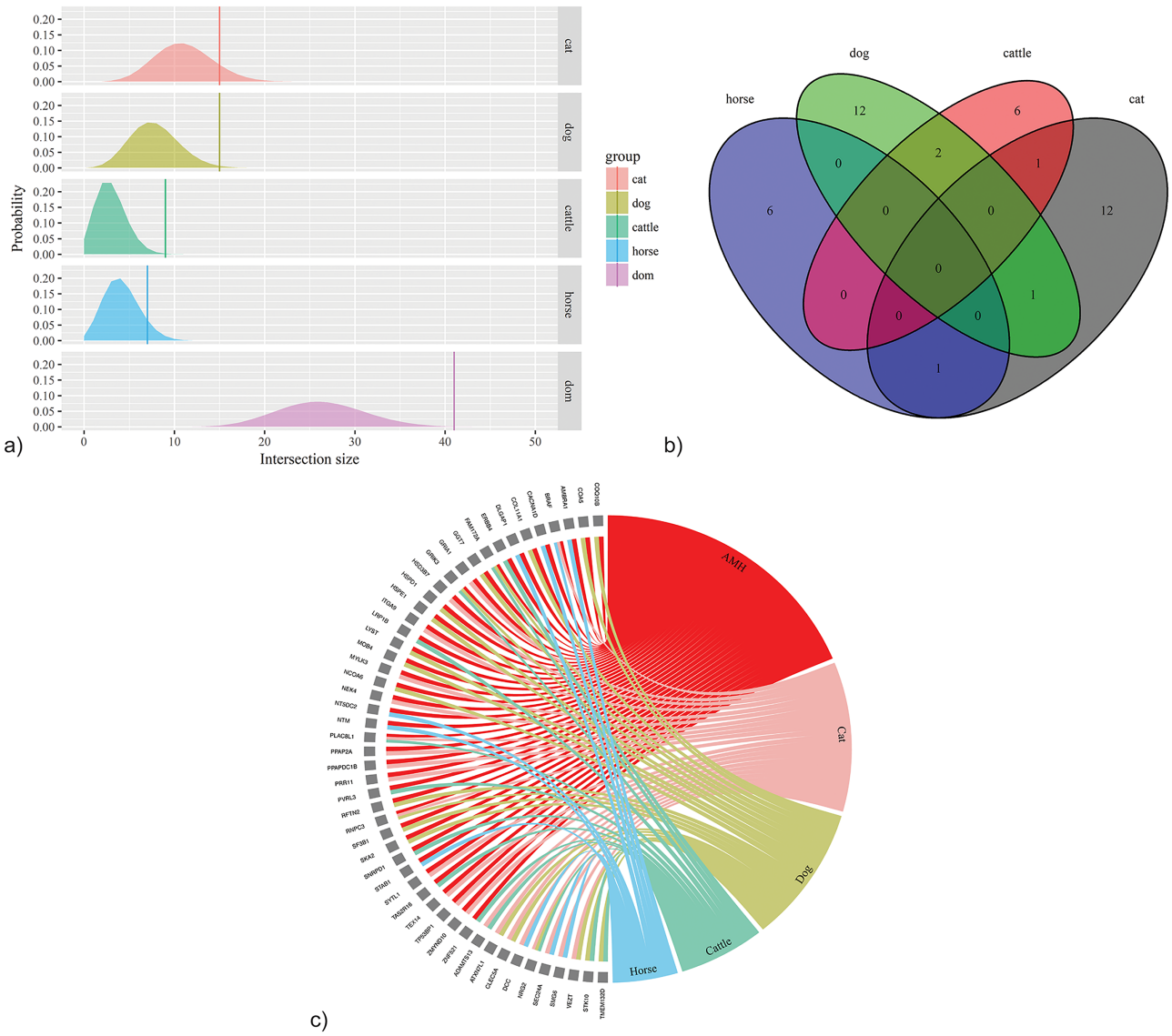


Fig 2. Graphical representations of overlapping genes showing signatures of positive selection in AMH and domesticated species. (a) Hypergeometric distributions for each group (individual domesticated species and the domesticate pool) with the probability of the intersection size found with AMH (v). cat: $v = 15$, $p = 0.1454$; dog: $v = 15$, $p = 0.0293$; cattle: $v = 9$, $p = 0.0028$; horse: $v = 7$, $p = 0.122$; dom: $v = 41$, $p = 0.0034$ (see S4 Table for details). (b) Venn diagram with the number of genes with signatures of positive selection overlapping between AMH and domesticated species. The number in each (sub)set is the number of genes showing signatures of positive selection shared by AMH and the respective species (see Table 1 and S2 Table for details). (c) Graph displaying the overlapping genes showing evidence of positive selection in AMH and one or more domesticated species ($n = 41$), and genes with evidence of positive selection in two or more domesticates (but not AMH) ($n = 9$) (see S1–S3 Tables for details).

https://doi.org/10.1371/journal.pone.0185306.g002

(the other is a gene of unknown function) [52]. Mutations in *RNPC3* cause growth hormone deficiencies in humans resulting from pituitary hypoplasia [62, 63]. In a similar vein, a gene showing an AMH-specific amino acid change and associated with a strong positive selection signal in AMH and in dogs, *NCOA6*, is a nuclear receptor coactivator that directly binds nuclear receptors and stimulates the transcriptional activities in a hormone-dependent fashion.

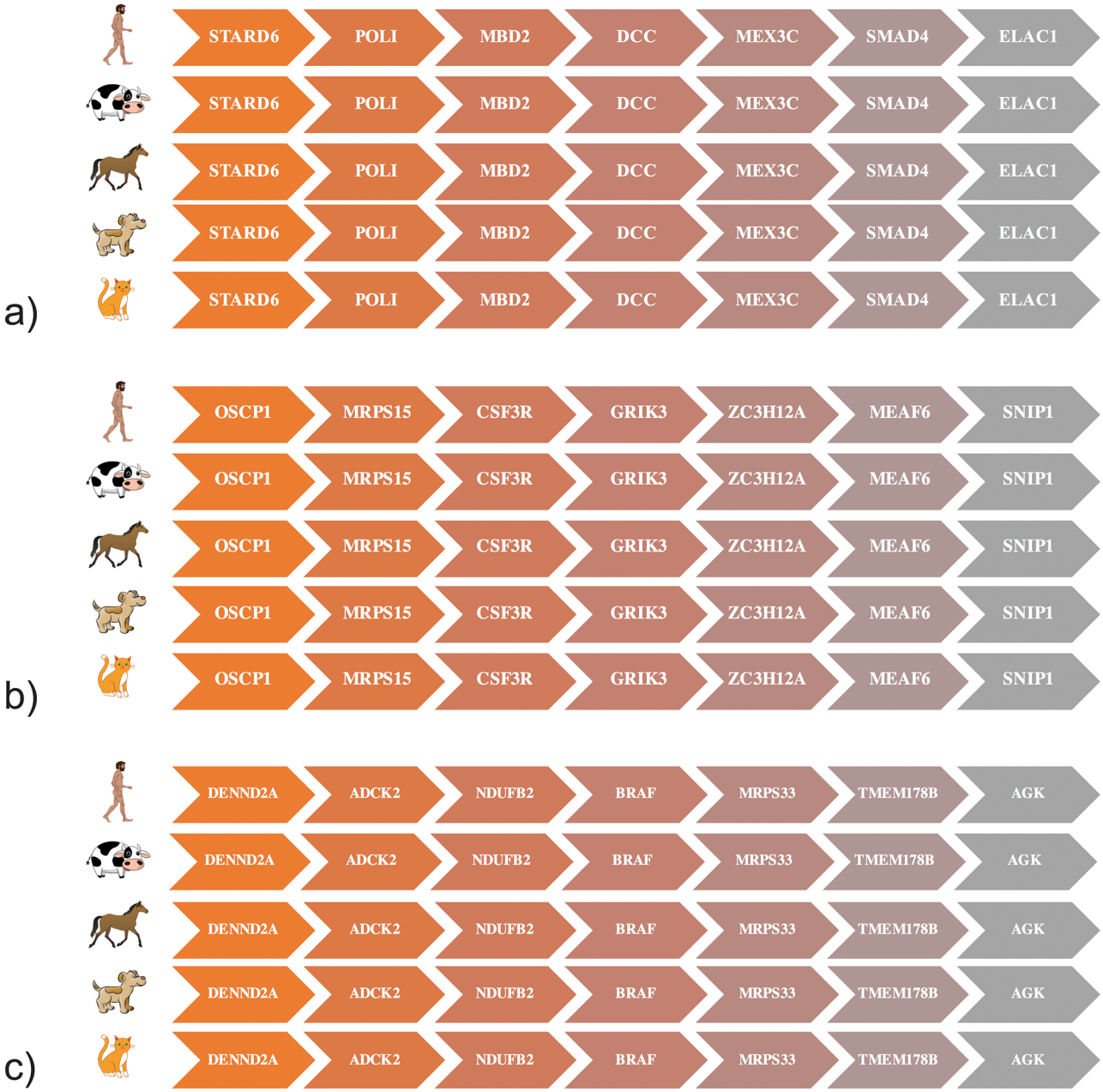


Fig 3. Examples of synteny analysis for 3 genes showing signatures of positive selection in AMH and domesticated species. Genes of interest (*DCC, GRIK3* and *BRAF*) and their 3 flanking protein-coding genes are shown in AMH, cattle, horse, dog and cat, illustrating their conserved syntenies. For other genes, see S6 Table.

https://doi.org/10.1371/journal.pone.0185306.g003

FAM172A, selected for in dogs, cattle, and AMH, may perhaps be worthy of note given its position on chromosome 5 neighboring *NR2F1*, which plays a role in regulating neural crest specifier genes and has undergone selection in AMH [64, 65]; The functionally related nuclear receptor *NR2F2* is involved in regulating embryonic stem cell differentiation [66] and implicated in neural crest development, and has been under selection in the domesticated fox [67].

PLAC8L1 is associated with positive selection signals in cats and cattle as well as in AMH, but there is only sparse evidence concerning its function. An autistic patient has been noted as having a microdeletion at chromosome 5q32, a location which includes *PLAC8L1* [68].

The two remaining genes in S3 Table, *BRAF* and *GRIK3*, deserve special attention. They are addressed in turn below.

ERK pathway

BRAF, under selection in the cat, horse, and AMH, is an important member of the ERK/MAPK signaling pathway, which has been shown to play a key role in synaptic plasticity, memory, and learning [69], and which, when disrupted, can lead to a broad range of syndromes comprising craniofacial defects and cognitive deficits [70]. *BRAF* is upstream of *ERK2*, which plays a critical role in neural crest development [71] and regulates neuronal gene expression in both the neocortex and hippocampus [69]. Both *BRAF* and *ERK2* inactivation can bring about syndromic symptoms by disrupting neural crest development [71]. *BRAF* is implicated in Noonan, Leopard, and Cardiofaciocutaneous syndromes, typical symptoms of which include prominent forehead, bitemporal narrowing, hypertelorism, and short stature, among other skeletal, cardiac, and craniofacial anomalies, frequently accompanied by moderate to severe mental retardation [72, 73]. *BRAF* interacts with other domestication-related genes, including *YWHAH* (under selection in the dog), *PPP2CA* (a neural crest-related gene, under selection in the horse), and *HER4/ERBB4*, another neural crest-related gene associated with a positive selection signal in cattle and AMH. Upstream of *BRAF*, *SOS1*, under selection in domesticated foxes, affects MAPK signaling, bringing about Noonan phenotypes [74]. Noonan syndrome-like phenotypes are associated with several genes that appear to have undergone selective sweeps in AMH. For instance, *CBL* is located in a region showing signals of a strong selective sweep in AMH compared to Altai Neanderthals [44], and, when mutated, has been shown to give rise to a Noonan syndrome-like disorder [75].

As mentioned above, the neuregulin (NRG) receptor *ERBB4*, which shows evidence of selection in humans, is part of the ERK/MAPK pathway and negatively regulates ERK via upstream phosphorylation of Raf-1 [76]. Loss of *Erbb4* function in mice has been shown to cause defects in hindbrain cranial neural crest cell pathfinding, including a caudal elongation of the trigeminal and geniculate ganglia [77]. This suggests a plausible role for *ERBB4* in preventing caudal extension in the derived AMH skull. *ERBB4* is one of many neural crest-related genes associated with selective signals in AMH (e.g., *SNAI2* [44], *CITED2* [44], *PRDM10* [46, 78], and others [38]), some of which show fixed or nearly fixed amino acid changes compared to Neanderthals. In addition, *NRG2* was the only gene that was found to be under selection in three of the four domesticated species in our study: cat, cattle, and dog. *NRG4* shows evidence of selection in cattle, and *NRG3*, in AMH. Incidentally, *NRG3* copy number and single nucleotide variants have been associated with Hirschsprung disease [79, 80]. This disease is very relevant in the context of domestication, as it affects the neural crest, associated with domestication syndrome [17, 38]. Quite a few genes associated with selective sweeps in AMH examined here (among them, *RET, ZEB2*, and *SLIT2*) have been linked to the disease [81, 82].

Enhanced ERBB4 signaling has been implicated in Angelman syndrome, an autism spectrum disorder marked by behavioral traits such as increased desire for social interaction, developmental delay, severe speech impairment, and a happy demeanour, although aggressive behavior has sometimes been reported [83–85]. Angelman-syndrome-like phenotypes are frequently associated with genes investigated here. One such observation concerns "the most intriguing variant fixed between dogs and wolves" [52], which is found in the 3'-UTR of *SLC9A6*. This gene encodes sodium/hydrogen exchanger protein 6, which is part of a network

related to the plasticity of glutaminergic neurons [86]. Cagan and Blass [52] note that loss-of-function mutations in this gene in humans can lead to Christianson syndrome, also known as "Angelman-like syndrome". Phenotypes typical of these patients include cognitive developmental delays, absence of speech, stereotyped repetitive hand movements, and postnatal microcephaly with a narrow face. Christianson syndrome is frequently characterized by a happy disposition with easily provoked laughter and smiling, an open mouth with excessive drooling and frequent visual fixation on hands. Several of these phenotypes resemble those that distinguish dogs from wolves.

We used Ingenuity Pathway Analysis software (QIAGEN, Redwood City, CA) to perform pathway analyses on the lists of genes in S1 and S2 Tables, as well as on the list of genes with amino acid replacement substitutions fixed in AMH and absent in archaic humans [87]. These analyses involved mining a database of literature on known interactions between the genes in each of these sets, and revealed that ERKs are among the most significant downstream targets of the interacting selected genes in each (see S1 Fig). Incidentally, an analysis of a domesticated pig (*Sus scrofa domesticus*) genome [88] suggests that the involvement of ERK pathway in domestication extends beyond the species we focused on in this study.

As a final note on the ERK pathway, we would like to highlight the presence of *CACNA1D* in S2 Table *CACNA1D* is a neural cell adhesion molecule that contributes to cell migration via activation of MAPK/ERK signaling [89]. This gene is one of several axon-guidance molecules we identified in our study. It is highly expressed in the adrenal glands [90], and, when mutated, gives rise to cerebral palsy/motor disorders [91]. It has been linked to auditory processing [92], and said to be among the positively selected genes in some vocal learners [93].

Glutamate receptors

The glutamate receptor *GRIK3* has been associated with positive selection signals in AMH, dog, and cattle, and interacts with other glutamate receptors associated with positive selection signals in the horse (*GRID1*) and cat (*GRIA1/2*). Polymorphisms in *GRIK3* and *GRID1* have been implicated in schizophrenia [94, 95], and *GRID1* neighbors *NRG3* (discussed above) at the schizophrenia susceptibility loci 10q22-q23 [96]. Developmental delays and craniofacial anomalies associated with a loss of genetic material at the *NRG3* locus, accompanied by a gain of material at the *DLGAP1* site, have also been reported [97]. *DLGAP1*, a scaffold-protein-coding gene at the postsynaptic density, is under selection in AMH and in the horse. This gene has been implicated in obsessive-compulsive disorders and interacts significantly with Shank proteins, mutations in which have been linked to autism spectrum disorders with impaired social interaction and communication [98–100]. DLGAP1 interacts with the glutamate receptor GRIK2, also implicated in obsessive-compulsive disorders [101].

Previous work by Li et al. [102] already pointed out that genes involved in glutamate metabolism show the greatest population differentiation by whole-genome comparison of dogs and wolves. Although such changes may be implicated in fear response differences between the dog and the wolf populations, Li et al. argue for a role in increasing excitatory synaptic plasticity in dogs rather than reducing fear response. As they point out, changes related to synaptic plasticity may have a significant impact on learning and memory. This is certainly true for cognitive specializations in humans, like language, since glutamate receptors have been shown to be differentially regulated in brain regions associated with vocal learning [103].

Genes under selection in multiple domesticates but not AMH

It is worth considering those genes under selection across domesticates, independently of their selection in AMH, for different reasons. First, our aim here is to explore the extent to which

self-domesticating processes in humans may have contributed to our species' anatomical, cognitive, and behavioral make-up. Uncovering genes of interest for (albeit often different) domesticating processes in other well-studied domesticated species is a promising way to pursue this goal. The most interesting genes should be those that are associated with positive selection signals across different species. Those genes under selection only in certain domesticates, but which strongly interact with genes under selection in other domesticated species, may prove central to a relevant domesticating process, given that these interactions may shed special light on relevant phenotypic traits. Similarly, certain genes associated with positive selection across different domesticates may have strong interactions with other genes that are under selection in AMH. We wish to highlight some of these here (for a full list, see S1 Table).

DCC (*DCC Netrin 1 receptor*), an axon-guidance mediator and neural crest-related gene, which shows signatures of positive selection in both the horse and the cat, interacts strongly with *DSCAM* (*Down Syndrome cell adhesion molecule*), another axon-guidance and neural crest-related gene, selected for in cattle. Of key significance is the interaction of *DCC* with the Slit/Robo pathway, especially given the proposed involvement of this pathway in vocal learning [104, 105] and the selection of both *ROBO2* and *SLIT2* in AMH [46]. The related gene *ROBO1* also shows evidence of selection in cattle. ROBO silences the attractive effect that Netrin 1 has on DCC, allowing SLIT2 to bind to this ligand and enabling axon pathfinding in the developing brain [106, 107]. *DCC* is involved in the organization of dopaminergic circuits within the cortex [108], and several association studies have identified *DCC* as a promising candidate for schizophrenia [109]. Importantly, an AMH-specific hCONDEL exists in a region upstream of *DCC*, although it is shared with Neanderthals [110]. However, a detailed examination of this gene on both the modern and archaic lines, reveals an accummulation of changes on this gene in AMH. In addition, several genes showing AMH-specific amino-acid substitutions, such as *NOVA1* and *RASA1*, both involved in neuronal development, are known to interact with *DCC* [111, 112], and could regulate it in a species-specific fashion. *RASA1* is associated with a strong selective signal in AMH, and has been shown to mediate Netrin 1-induced cortical axon outgrowth and guidance [112]. Together with the glutamate receptor changes discussed above, such modifications may have played an important role in generating aspects of the cognitive profile associated with modern humans, including a full-fledged language-ready brain.

We found several collagen-type genes with signatures of selection across domesticates. *COL22A1*, a gene under selection in the horse, significantly interacts with various similar genes associated with positive selection signals in other domesticates, particularly in the cat, including *COL11A1*, under selection in the dog and AMH. *COL22A1* and *COL11A1* exhibit increased expression in the bone tissue and hippocampus of mice with some of the symptoms of Kleefstra Syndrome (developmental delay, hypotonia, and craniofacial abnormalities), which is often accompanied by autistic symptoms and intellectual disability in humans [113, 114].

Archaic-derived alleles

To the best of our knowledge, no comprehensive selective sweep analysis exists for Neanderthals. We examined the genes associated with archaic-derived alleles [115] and found that no genes in S1 Table display reported archaic-derived alleles. While this could be due to the modest number of archaic-specific SNCs known at the time of writing, we find this to be an important contrast with the situation that obtains with AMH, in light of the self-domestication hypothesis. It is striking that Castellano et al. [115] highlight genes involved in skeletal development and associated with aggressive phenotypes in their comparison of archaic *Homo* and AMH.

We also examined data concerning nearly fixed ancestral or derived SNPs in archaic lineages that crop up as variants in modern-day populations. Despite the many confounding

factors as to how the relevant mutated genes might interact in different genetic contexts, one might still expect certain archaic-selected SNPs to exhibit somewhat 'underdomesticated' phenotypes when occurring as AMH variants. In this sense, mutations imitating ancestral SNPs found in archaic lineages may be able to tell us a great deal about the evolution of our lineage, by allowing us to glimpse some aspects of the ancestral genotype. Among those mutations we found through an exhaustive literature review, there is an ancestral S330A mutation of *SLITRK1* that may be involved in obsessive-compulsive disorders like Tourette's Syndrome [116, 117]. Different amino acid changes around the site of an AMH-specific derived protein, ADSL (A429V), can bring about adenylosuccinate lyase deficiency (R426H; D430N [118]), the symptoms of which include developmental delay, autistic-like traits, aggressiveness, and microcephaly [119].

Discussion

As already mentioned in the Introduction, several scholars have pointed out that there are several routes to domestication. We should therefore expect genes targeted by domestication processes to differ considerably across species. Nevertheless, reviewing the molecular events associated with domestication reveals common themes, with significant numbers of genes related to brain function and behavior, anatomy, and diet, across domesticates. This is consistent with the view that domestication may be best represented as a spectrum or continuum [120], with a polygenic basis and non-uniform symptomatology. This state of affairs is reflected in significant brain gene expression differences across domesticates, with the majority of these changes being species-specific [121].

Because of these findings, we find the overlaps listed in S1 and S2 Tables and the associated functions and pathways discussed in the Results section all the more relevant, especially because they converge to a large extent with what is to be expected from the neural crest-based hypothesis [17] put forth to capture the common mechanistic basis of domestication events. A disruption in neural crest developmental programs might be the source of changes spanning multiple organ systems and morphological structures [17], and the genes examined here seem to broadly support this view. It is quite possible that a neural crest-based explanation won't apply to all domesticates [16], but it is interesting that this hypothesis finds its strongest support in species like dogs (see also [122]), which have been argued to be self-domesticated [34]. Recall that the goal of the present study was not to provide molecular evidence for a general theory of domestication, but rather to identify domestication-related pathways that could be suggestive of a self-domestication process in AMH. The fact that we find neural crest-related changes in AMH compared to Neanderthals/Denisovans, and that such changes are also found in another species hypothesized to have undergone a self-domestication process, reinforces our hypothesis that self-domestication took place in our species.

Apart from neural crest-related genes and pathways, we identified common themes pertaining to neuronal development, synaptic plasticity, and enhanced learning. These categories are often mentioned in studies on selective sweeps in AMH (e.g., [46]). These results are in line with claims in other studies on domestication [49, 123–125], where categories like 'neurological process' frequently stand out strongly in gene ontology category enrichment analyses. This potentially lends credence to claims pairing domestication and a certain type of intelligence [126]. It is also not unreasonable to suspect that byproducts of the domestication process, such as enhanced sensory-motor perceptual and learning pathways, may provide a foundation for more complex communicative abilities, including vocal learning abilities [39, 127].

In a similar vein, among the genes under selection in both AMH and one or more domesticates, as well as in those under selection in multiple domesticates though not AMH, one finds multiple strong candidates for neurodevelopmental diseases and syndromes (see also [128]). This could be seen as an additional piece of evidence suggestive of a self-domestication process in AMH. A build-up of deleterious alleles is documented across domesticated species when compared to their wild counterparts. For instance, there is a higher frequency of non-synonymous substitutions in the nuclear DNA of domesticated dogs relative to gray wolves [129], and the same is true of their mitochondrial DNA [130]. A higher frequency of non-synonymous substitutions in domesticated yaks compared to the wild yaks has also been reported [131]. This build-up of deleterious alleles has been described as the 'cost of domestication' [132], which, if true, could be a byproduct of self-domestication in AMH, too.

A study like the present one suffers from several limitations. While we have tried to make our comparisons as fair as possible, we have relied on genomic data that necessarily reflect the current state of the art for the various species we examined. The lists of genes associated with signals of positive selection are derived from the literature, and were generated using different analytical tools. While we have done our best to minimize the number of simplifying assumptions (see Methods), we must point out that even within a single species (e.g., AMH), no two studies completely agree on a definitive list. Indeed, in some cases, they produce lists of very different sizes. In addition, we may have missed important genes of interest due to the lack of information on them in the various databases we consulted. While it is to be hoped that some of these limitations will be overcome in the future, we think that the overlaps discussed in this study should encourage further detailed examination of these genes and the processes in which they take part. Last, but not least, it remains to be determined experimentally that the overlaps discovered here are indeed associated with mutations that led to similar functional effects across species.

We could have been more strict about our notion of convergence, and restrict our attention to genes where the exact same difference (e.g., the same amino acid substitution) could be detected across species (for an early attempt along these lines, see [133]). But given that convergent evolution is often hypothesized to occur in the absence of this very strict notion of convergence—for instance, convergent evolution in the domain of vocal learning is related to non-identical changes in *FOXP2* across vocal learners [134]—we feel justified in our approach.

Methods

Data

To identify signatures of a self-domestication process in AMH, we first constructed a list of genes associated with signs of positive selection in AMH compared to Neanderthals and Denisovans, which yielded a total of 742 genes. We then compared this list to the genes independently argued to be associated with positive selection in domesticated species versus their wild counterparts, which numbered 691 in total, and examined the overlap between these two gene lists.

For AMH-Neanderthal/Denisovan comparisons, we made use of findings based on high-quality genome reconstructions, specifically: the list of genes in regions of putative selective sweeps, together with pathway and disease annotation, of Prüfer et al. [44]; the list of genes from the top 20 candidate regions for the modern human ancestral branch in the work of Racimo [48]; and the extended list of genomic regions predicted to underlie positively selected human specific traits by Peyrégne et al. [46].

We included in our study a range of domesticated species for which detailed genetic information is available. These species offer representative examples of the various routes to

domestication [11], as well as different temporal windows for domestication. The species include: dog (*Canis familiaris*) [49, 50, 52], cat (*Felis catus*) [47], horse (*Equus caballus*) [45], and taurine cattle (*Bos taurus*) [51]. We homogenized the nomenclature across gene sets as best we could.

We also examined other species, including the rabbit (*Oryctolagus cuniculus*) [125], and bonobo (*Pan paniscus*) [135]. In the end, the lists of genes under selection for these species (compared to their wild counterparts) were too small to draw any firm conclusions.

To help us understand domestication-related changes better, we made use of the comparison of two lines of rats (*Rattus norvegicus*) selected for tame and aggressive behaviour to identify genetic loci that differ between the lines [136], the comparison of gene expression levels in the brains of domesticated and wild animals [121], genomic signatures of domestication in neurogenetic genes in *Drosophila melanogaster* (in which neurogenetic genes have been claimed to be associated with signs of positive selection [123]), and the genetic divergence between foxes (*Vulpes vulpes*) that were selected for tame and aggressive behavior [67].

For the Great Ape comparison—chimpanzee (*Pan t. troglodytes*), orangutan (*Pongo abelii*), and gorilla (*G. g. gorilla*)—we made use of positive and balancing selection and selective sweep data from Cagan et al. [137] (Tables S6, S18(68), S19(69), S20(70), S24(74), and S97).

For AMH comparisons with the Eurasian wolf (*Canis lupus lupus*) we used data from Stronen et al. [54] (Tables 2, S3, and S5: genes under selection associated with environmental and geographic variables or with no obvious spatial patterns) and Pilot et al. [53] (Table S4: genes adjacent to loci putatively under selection in European wolves). For the wisent (*Bison bonasus*) we used data from Gautier et al. [55] (Table S3: genes under positive selection between the wisent and bovine lineages) and Wang et al. [56] (Table S14: genes under positive selection in the wisent).

Methods

In order to test the significance of the overlap between domestication-related genes and genes showing signals of positive selection and selective sweep in AMH, a hypergeometric intersection test was performed using the *R* software [138] and the *R* package `hint` [139]. A hypergeometric intersection distribution can be employed to compute the probability of picking an intersection of size v when drawing independently and without replacement from two sets A and B composed of objects of n categories, with a and b number of draws, respectively (where $a \neq b$) [139].

As a model of our data we chose as a simplifying assumption $n = 19{,}500$ as the average number of protein-coding genes for all the species taken into consideration. From the original lists, we removed antisense RNA genes (non coding), miRNAs, and other non-coding transcripts/products listed in the original tables.

From this modeled genome, a total of $a = 691$ genes were drawn from the domesticate pool (comprising cat, dog, cattle, and horse), while $b = 742$ genes were drawn from the total AMH pool. The resulting intersection size (i.e., the number of genes associated with positive selection signals both in AMH and in one or more domesticate) was $v = 41$. The `hint.test` function was then employed to test the significance of this intersection, obtaining $p < 0.01$.

A Monte Carlo simulation was performed using Matlab (MathWorks, Natick, MA) to confirm these results. Two random samples, of lengths 691 and 742 (with no replacement), were drawn from a pool representing 19,500 genes using Matlab's random number generation function. These simulated draws were performed 1,000,000 times and the percentage of trials in which the intersection was ≥ 41 was calculated. The results revealed that 0.33% of trials had intersections of this size.

Since we pooled data for positive selection and selective sweep in AMH from different sources, hypergeometric intersection tests were carried out between the domestication pool and the pool of each AMH dataset used in this study. A significant intersection was found with the data in [44] ($a = 691$, $b = 108$, $v = 9$; $p < 0.05$) and with the combined data from [44] and [48] ($a = 691$, $b = 419$, $v = 24$, $p < 0.05$).

Overlaps with domesticates were tested for Great Apes, using data from Cagan et al. [137]. For chimpanzee (*Pan t. troglodytes*), $b = 415$ with $v = 16$; for orangutan (*Pongo abelii*), $b = 500$ with $v = 20$; for gorilla (*G. g. gorilla*), $b = 426$ with $v = 12$. The hypergeometric intersection tests yielded non-significant results for all these intersections. Monte Carlo simulations, performed as described above, *mutatis mutandis*, showed that intersections of these sizes occurred in a large fraction of trials (40.11% of trials for chimpanzee; 32% for orangutan; 82.89% for gorilla). As in the case of AMH, overlaps with individual domesticates were tested, with no significant results.

We tested overlaps with the Eurasian wolf (*Canis lupus lupus*) using data from Stronen et al. [54] (Table 2: $b = 32$ with $v = 3$, S3: $b = 70$ with $v = 0$, S5: $b = 33$ with $v = 1$) and [140] ($b = 32$ with $v = 1$). For the wisent (*Bison bonasus*) we tested overlaps using data from Gautier et al. [55] ($b = 425$ with $v = 11$) and Wang et al. [56] ($b = 72$ with $v = 3$). None of the overlaps between these non-domesticated species and AMH were significant.

For synteny analysis, we used the genomic data available for each species in the NCBI (https://www.ncbi.nlm.nih.gov/) and Ensemble (http://www.ensembl.org/index.html) databases. In S1 Table, for each of the 41 overlapping genes we included the 4 protein-coding genes flanking the region of interest. We added more flanking protein-coding genes only in the instances where some event (e.g., gene insertion or local duplication) rendered the synteny less clear. We also used NCBI Gene Search and BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to confirm that some of the genes surrounding the genes of interest were the same genes across taxa, with different names in some species' assemblies.

We then examined the functions of the genes in S1, S2 and S3 Tables, paying close attention to the pathways in which they are involved, and to their interactions with other genes already highlighted in the domestication literature. In addition to performing an exhaustive PubMed (http://www.ncbi.nlm.nih.gov/pubmed) search on each of the genes, we drew upon the information available in Genecards (http://genecards.org), Uniprot (http://www.uniprot.org/), String 10.0 (http://string-db.org), and Biogrid 3.4 (http://thebiogrid.org) to identify potential protein-protein interactions and Gene Ontology category enrichment signals. Additionally, we fed the gene lists in S1 and S2 Tables into Ingenuity Pathway Analysis software (QIAGEN, Redwood City, CA) and used the Core Analysis tools to study the associated gene networks and functions. The two major networks generated by these analyses, in which the centrality of the ERK pathway is visible, are provided S1 Fig.

Furthermore, we gathered information about the expression patterns of these genes, concentrating on those genes with relatively high expression in tissues such as brain, bone, and adrenal glands. For this, we relied on the following resources: Brainspan (http://www.brainspan.org), Human Brain Transcriptome (http://hbatlas.org), Bgee (http://bgee.org), Proteomics DB (https://proteomicsdb.org), Human Protein Atlas (http://www.proteinatlas.org), Gene Enrichment Profiler (http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/index.html), and GTex (http://www.gtexportal.org). For the information presented in the Supplementary Material, we consulted the following databases: KEGG Pathways and Disease (http://www.kegg.jp/kegg/), PANTHER (http://www.pantherdb.org), Reactome Pathway Database (http://www.reactome.org), OMIM (http://omim.org), and MalaCards (http://www.malacards.org/).

Supporting information

S1 Fig. Networks of overlapping genes (a) between domesticated species and (b) between AMH and domesticated species, generated using QIAGEN Ingenuity Pathway Analysis software.
(PDF)

S1 Table. Genes overlapping between at least two domesticated species.
(PDF)

S2 Table. Genes overlapping between AMH and domesticated species.
(PDF)

S3 Table. Genes overlapping between AMH and two or more domesticated species.
(PDF)

S4 Table. Gene lists from the sources for AMH, domesticated species, and great apes. Gene lists and statistical analysis of data from domesticates, between AMH and each domesticate, and between great apes and domesticates.
(PDF)

S5 Table. Genes overlapping between AMH, non-domesticated Canis (grey wolf), and non-domesticated bovine (wisent, European bison). Gene lists and statistical analysis of data from AMH and non-domesticated species.
(PDF)

S6 Table. Synteny analysis of the genes overlapping between AMH and domesticated species, and those under selection in multiple domesticated species.
(PDF)

Acknowledgments

We acknowledge the help of Nina Riddell in generating graph (c) in Fig 2. The *Homo* skull drawings in Fig 1 were done by Francisco Peña, who granted us permission to use them here. The wolf and dog skulls in Fig 1 were adapted from `DigiMorph.org`, with kind permission from Dr. Timothy Rowe. We thank Kay Sušelj for assistance with the Monte Carlo simulations.

Author Contributions

Conceptualization: Constantina Theofanopoulou, Simone Gastaldon, Cedric Boeckx.

Data curation: Constantina Theofanopoulou, Simone Gastaldon, Thomas O'Rourke, Angela Messner, Pedro Tiago Martins, Saleh Alamri.

Formal analysis: Constantina Theofanopoulou, Simone Gastaldon, Bridget D. Samuels, Angela Messner, Francesco Delogu, Saleh Alamri.

Funding acquisition: Cedric Boeckx.

Methodology: Simone Gastaldon, Bridget D. Samuels, Pedro Tiago Martins, Francesco Delogu, Cedric Boeckx.

Supervision: Cedric Boeckx.

Visualization: Bridget D. Samuels, Pedro Tiago Martins.

Writing – original draft: Constantina Theofanopoulou, Simone Gastaldon, Thomas O'Rourke, Cedric Boeckx.

Writing – review & editing: Constantina Theofanopoulou, Simone Gastaldon, Thomas O'Rourke, Bridget D. Samuels, Angela Messner, Pedro Tiago Martins, Saleh Alamri, Cedric Boeckx.

References

1. Leach H. Human domestication reconsidered. Current Anthropology. 2003; 44(3):349–368. https://doi.org/10.1086/368119
2. Thomas J. Self-domestication and language evolution. University of Edinburgh. test; 2014.
3. Darwin C. The descent of man and selection in relation to sex. vol. 1. Murray; 1888.
4. Boas F. The mind of primitive man (Rev. ed.). New York. 1938; p. 122–44.
5. Morey DF. The early evolution of the domestic dog. American Scientist. 1994; 82(4):336–347.
6. O'Connor TP. Working at relationships: another look at animal domestication. Antiquity. 1997; 71 (271):149–156. https://doi.org/10.1017/S0003598X00084635
7. Trut L. Early Canid Domestication: The Farm-Fox Experiment Foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. American Scientist. 1999; 87(2):160–169.
8. Belyaev DK, et al. Destabilizing selection as a factor in domestication. Journal of Heredity. 1979; 70 (5):301–308. https://doi.org/10.1093/oxfordjournals.jhered.a109263 PMID: 528781
9. Zeder MA. Central questions in the domestication of plants and animals. Evolutionary Anthropology: Issues, News, and Reviews. 2006; 15(3):105–117. https://doi.org/10.1002/evan.20101
10. Zeder MA, Emshwiller E, Smith BD, Bradley DG. Documenting domestication: the intersection of genetics and archaeology. TRENDS in Genetics. 2006; 22(3):139–155. https://doi.org/10.1016/j.tig.2006.01.007 PMID: 16458995
11. Zeder MA. Pathways to animal domestication. Biodiversity in agriculture: Domestication, evolution and sustainability. 2012; p. 227–259.
12. Marshall-Pescini S, Cafazzo S, Virányi Z, Range F. Integrating social ecology in explanations of wolf–dog behavioral differences. Current Opinion in Behavioral Sciences. 2017; 16:80–86. https://doi.org/10.1016/j.cobeha.2017.05.002
13. Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, et al. The Near Eastern origin of cat domestication. Science. 2007; 317(5837):519–523. https://doi.org/10.1126/science.1139518 PMID: 17600185
14. Driscoll CA, Macdonald DW, O'Brien SJ. From wild animals to domestic pets, an evolutionary view of domestication. Proceedings of the National Academy of Sciences. 2009; 106(Supplement 1):9971–9978. https://doi.org/10.1073/pnas.0901586106
15. Budiansky S. The covenant of the wild: why animals chose domestication: with a new preface. Yale University Press; 1992.
16. Sánchez-Villagra MR, Geiger M, Schneider RA. The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. Open Science. 2016; 3(6):160107.
17. Wilkins AS, Wrangham RW, Fitch WT. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. Genetics. 2014; 197(3):795–808. https://doi.org/10.1534/genetics.114.165423 PMID: 25024034
18. Hublin JJ, Neubauer S, Gunz P. Brain ontogeny and life history in Pleistocene hominins. Phil Trans R Soc B. 2015; 370(1663):20140062. https://doi.org/10.1098/rstb.2014.0062 PMID: 25602066
19. Lacruz R, Bromage T, O'Higgins P, Arsuaga JL, Stringer C, Godinho R, et al. Facial ontogeny in Neanderthals and their ancestors. Nature Communications. 2015;.
20. Maureille B, Bar D. The premaxilla in Neandertal and early modern children: ontogeny and morphology. Journal of human evolution. 1999; 37(2):137–152. https://doi.org/10.1006/jhev.1999.0312 PMID: 10444349
21. Cieri RL, Churchill SE, Franciscus RG, Tan J, Hare B. Craniofacial feminization, social tolerance, and the origins of behavioral modernity. Current Anthropology. 2014; 55(4):419–443. https://doi.org/10.1086/677209

22. Zilberman U, Smith P. A comparison of tooth structure in Neanderthals and early Homo sapiens sapiens: a radiographic study. Journal of anatomy. 1992; 180(Pt 3):387. PMID: 1487432
23. Carter K, Worthington S. Morphologic and Demographic Predictors of Third Molar Agenesis A Systematic Review and Meta-analysis. Journal of dental research. 2015; 94(7):886–894. https://doi.org/10.1177/0022034515581644 PMID: 25883107
24. de León MSP, Golovanova L, Doronichev V, Romanova G, Akazawa T, Kondo O, et al. Neanderthal brain size at birth provides insights into the evolution of human life history. Proceedings of the National Academy of Sciences. 2008; 105(37):13764–13768. https://doi.org/10.1073/pnas.0803917105
25. Hare B, Wobber V, Wrangham R. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. Animal Behaviour. 2012; 83(3):573–585. https://doi.org/10.1016/j.anbehav.2011.12.007
26. Trut L, Oskina I, Kharlamova A. Animal evolution during domestication: the domesticated fox as a model. Bioessays. 2009; 31(3):349–360. https://doi.org/10.1002/bies.200800070 PMID: 19260016
27. Künzl C, Sachser N. The behavioral endocrinology of domestication: a comparison between the domestic guinea pig (Cavia apereaf. porcellus) and its wild ancestor, the cavy (Cavia aperea). Hormones and Behavior. 1999; 35(1):28–37. https://doi.org/10.1006/hbeh.1998.1493 PMID: 10049600
28. Schaefer K, Fink B, Mitteroecker P, Neave N, Bookstein FL. Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. Collegium antropologicum. 2005; 29(2):415–419. PMID: 16417137
29. Nelson E, Rolian C, Cashmore L, Shultz S. Digit ratios predict polygyny in early apes, Ardipithecus, Neanderthals and early modern humans but not in Australopithecus. Proceedings of the Royal Society of London B: Biological Sciences. 2011; 278(1711):1556–1563. https://doi.org/10.1098/rspb.2010.1740
30. Knight C, Power C, Watts I. The human symbolic revolution: a Darwinian account. Cambridge archaeological journal. 1995; 5(01):75–114. https://doi.org/10.1017/S0959774300001190
31. Stringer C. The origin and evolution of Homo sapiens. Phil Trans R Soc B. 2016; 371 (1698):20150237. https://doi.org/10.1098/rstb.2015.0237 PMID: 27298468
32. Dugatkin LA, Trut L. How to Tame a Fox (and Build a Dog): Visionary Scientists and a Siberian Tale of Jump-Started Evolution. Chicago: University of Chicago Press; 2017.
33. Singh N, Albert FW, Plyusnina I, Trut L, Pääbo S, Harvati K. Facial shape differences between rats selected for tame and aggressive behaviors. PloS one. 2017; 12(4):e0175043. https://doi.org/10.1371/journal.pone.0175043 PMID: 28369080
34. Hare B. Survival of the Friendliest: Homo Sapiens Evolved via Selection for Prosociality. Annual Review of Psychology. 2016; 68(1). PMID: 27732802
35. Honda E, Okanoya K. Acoustical and syntactical comparisons between songs of the white-backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica). Zoological Science. 1999; 16(2):319–326. https://doi.org/10.2108/zsj.16.319
36. Okanoya K. The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Annals of the New York Academy of Sciences. 2004; 1016(1):724–735. https://doi.org/10.1196/annals.1298.026 PMID: 15313802
37. Okanoya K. Evolution of song complexity in Bengalese finches could mirror the emergence of human language. Journal of Ornithology. 2015; 156(1):65–72. https://doi.org/10.1007/s10336-015-1283-5
38. Benítez-Burraco A, Theofanopoulou C, Boeckx C. Globularization and domestication. Topoi. 2016; p. 1–14.
39. Boeckx C. Evolution of Language. In: Kaas J, editor. Evolution of Nervous Systems, 2nd ed., vol. 4. London: Elsevier; 2017. p. 325–339.
40. Okanoya K. Sexual communication and domestication may give rise to the signal complexity necessary for the emergence of language: An indication from songbird studies. Psychonomic Bulletin & Review. 2017; p. 1–5.
41. Kirby S. Culture and biology in the origins of linguistic structure. Psychonomic Bulletin & Review. 2017;. https://doi.org/10.3758/s13423-016-1166-7
42. Llamas B, Willerslev E, Orlando L. Human evolution: a tale from ancient genomes. Phil Trans R Soc B. 2017; 372(1713):20150484. https://doi.org/10.1098/rstb.2015.0484 PMID: 27994125
43. Orlando L. Ancient DNA and the Study of Animal Domestication. Annual Review of Animal Biosciences. 2016; 5(1). PMID: 27813680
44. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014; 505(7481):43–49. https://doi.org/10.1038/nature12886 PMID: 24352235

45. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proceedings of the National Academy of Sciences. 2014; 111(52):E5661–E5669. https://doi.org/10.1073/pnas.1416991111

46. Peyrégne S, Dannemann M, Prüfer K. Detecting ancient positive selection in humans using extended lineage sorting. bioRxiv. 2016; p. 092999.

47. Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SM, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. Proceedings of the National Academy of Sciences. 2014; 111(48):17230–17235. https://doi.org/10.1073/pnas.1410083111

48. Racimo F. Testing for ancient selection using cross-population allele frequency differentiation. Genetics. 2016; 202(2):733–750. https://doi.org/10.1534/genetics.115.178095 PMID: 26596347

49. Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, Han E, Davis BW, Gronau I, et al. Demographically-based evaluation of genomic regions under selection in domestic dogs. PLoS Genet. 2016; 12 (3):e1005851. https://doi.org/10.1371/journal.pgen.1005851 PMID: 26943675

50. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 2013; 495(7441):360–364. https://doi.org/10.1038/nature11837 PMID: 23354050

51. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 2014; 10(2):e1004148. https://doi.org/10.1371/journal.pgen.1004148 PMID: 24586189

52. Cagan A, Blass T. Identification of genomic variants putatively targeted by selection during dog domestication. BMC evolutionary biology. 2016; 16(1):1. https://doi.org/10.1186/s12862-015-0579-7

53. Pilot M, Greco C, et al. Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity. 2014; 112(4):428. https://doi.org/10.1038/hdy.2013.122 PMID: 24346500

54. Stronen AV, Jędrzejewska B, Pertoldi C, Demontis D, Randi E, Niedziałkowska M, et al. Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore. Ecology and evolution. 2015; 5(19):4410–4425. https://doi.org/10.1002/ece3.1695 PMID: 26664688

55. Gautier M, Moazami-Goudarzi K, Levéziel H, Parinello H, Grohs C, Rialle S, et al. Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. Molecular biology and evolution. 2016; 33(11):2801–2814. https://doi.org/10.1093/molbev/msw144 PMID: 27436010

56. Wang K, Wang L, Lenstra JA, Jian J, Yang Y, Hu Q, et al. The genome sequence of the wisent (Bison bonasus). GigaScience. 2017; 6(4):1–5.

57. Buntjer J, Otsen M, Nijman I, Kuiper M, Lenstra J. Phylogeny of bovine species based on AFLP fingerprinting. Heredity. 2002; 88(1):46. https://doi.org/10.1038/sj.hdy.6800007 PMID: 11813106

58. Bibi F. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evolutionary Biology. 2013; 13(1):166. https://doi.org/10.1186/1471-2148-13-166 PMID: 23927069

59. Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield D, et al. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science. 2013; 342 (6160):871–874. https://doi.org/10.1126/science.1243650 PMID: 24233726

60. Frantz LA, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science. 2016; 352(6290):1228–1231. https://doi.org/10.1126/science.aaf3161 PMID: 27257259

61. Wang GD, Zhai W, Yang HC, Wang L, Zhong L, Liu YH, et al. Out of southern East Asia: the natural history of domestic dogs across the world. Cell research. 2016; 26(1):21. https://doi.org/10.1038/cr.2015.147 PMID: 26667385

62. Argente J, Flores R, Gutiérrez-Arumí A, Verma B, Martos-Moreno GÁ, Cuscó I, et al. Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. EMBO molecular medicine. 2014; p. e201303573.

63. Gucev Z, Polenakovic M, Tasic V, LeBouc Y, Klammt J, Pfaeffle R, et al. Severe Isolated Growth Hormone Deficiency and Myopathy in Two Brothers with RNPC3 Mutation. ESPE Abstracts. 2015;.

64. Rada-Iglesias A, Bajpai R, Prescott S, Brugmann SA, Swigut T, Wysocka J. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell stem cell. 2012; 11(5):633–648. https://doi.org/10.1016/j.stem.2012.07.006 PMID: 22981823

65. Simões-Costa M, Bronner ME. Establishing neural crest identity: a gene regulatory recipe. Development. 2015; 142(2):242–257. https://doi.org/10.1242/dev.105445 PMID: 25564621

66. Rosa A, Brivanlou AH. A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation. The EMBO journal. 2011; 30 (2):237–248. https://doi.org/10.1038/emboj.2010.319 PMID: 21151097

67. Johnson JL, Wittgenstein H, Mitchell SE, Hyma KE, Temnykh SV, Kharlamova AV, et al. Genotyping-by-sequencing (GBS) detects genetic structure and confirms behavioral QTL in tame and aggressive foxes (Vulpes vulpes). PloS one. 2015; 10(6):e0127013. https://doi.org/10.1371/journal.pone.0127013 PMID: 26061395

68. Gau SSF, Liao HM, Hong CC, Chien WH, Chen CH. Identification of two inherited copy number variants in a male with autism supports two-hit and compound heterozygosity models of autism. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2012; 159(6):710–717. https://doi.org/10.1002/ajmg.b.32074

69. Sweatt JD. The neuronal MAP kinase cascade: a biochemical signal integration system subserving synaptic plasticity and memory. Journal of neurochemistry. 2001; 76(1):1–10. https://doi.org/10.1046/j.1471-4159.2001.00054.x PMID: 11145972

70. Cesarini L, Alfieri P, Pantaleoni F, Vasta I, Cerutti M, Petrangeli V, et al. Cognitive profile of disorders associated with dysregulation of the RAS/MAPK signaling cascade. American Journal of Medical Genetics Part A. 2009; 149(2):140–146. https://doi.org/10.1002/ajmg.a.32488

71. Newbern J, Zhong J, Wickramasinghe RS, Li X, Wu Y, Samuels I, et al. Mouse and human phenotypes indicate a critical conserved role for ERK2 signaling in neural crest development. Proceedings of the National Academy of Sciences. 2008; 105(44):17115–17120. https://doi.org/10.1073/pnas.0805239105

72. Keyte A, Hutson MR. The neural crest in cardiac congenital anomalies. Differentiation. 2012; 84 (1):25–40. https://doi.org/10.1016/j.diff.2012.04.005 PMID: 22595346

73. Sarkozy A, Carta C, Moretti S, Zampino G, Digilio MC, Pantaleoni F, et al. Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. Human mutation. 2009; 30(4):695–702. https://doi.org/10.1002/humu.20955 PMID: 19206169

74. Roberts AE, Araki T, Swanson KD, Montgomery KT, Schiripo TA, Joshi VA, et al. Germline gain-of-function mutations in SOS1 cause Noonan syndrome. Nature genetics. 2007; 39(1):70–74. https://doi.org/10.1038/ng1926 PMID: 17143285

75. Martinelli S, De Luca A, Stellacci E, Rossi C, Checquolo S, Lepri F, et al. Heterozygous germline mutations in the CBL tumor-suppressor gene cause a Noonan syndrome-like phenotype. The American Journal of Human Genetics. 2010; 87(2):250–257. https://doi.org/10.1016/j.ajhg.2010.06.015 PMID: 20619386

76. Hatakeyama M, Kimura S, Takashi N, Kawasaki T, Yumoto N, Ichikawa M, et al. A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. Biochemical Journal. 2003; 373(2):451–463. https://doi.org/10.1042/BJ20021824 PMID: 12691603

77. Golding JP, Trainor P, Krumlauf R, Gassmann M. Defects in pathfinding by cranial neural crest cells in mice lacking the neuregulin receptor ErbB4. Nature cell biology. 2000; 2(2):103–109. https://doi.org/10.1038/35000058 PMID: 10655590

78. Racimo F, Kuhlwilm M, Slatkin M. A test for ancient selective sweeps and an application to candidate sites in modern humans. Molecular biology and evolution. 2014; 31(12):3344–3358. https://doi.org/10.1093/molbev/msu255 PMID: 25172957

79. Tang CSM, Cheng G, So MT, Yip BHK, Miao XP, Wong EHM, et al. Genome-wide copy number analysis uncovers a new HSCR gene: *NRG3*. PLoS Genet. 2012; 8(5):10.1371. https://doi.org/10.1371/journal.pgen.1002687

80. Yang J, Duan S, Zhong R, Yin J, Pu J, Ke J, et al. Exome sequencing identified *NRG3* as a novel susceptible gene of Hirschsprung's disease in a Chinese population. Mol Neurobiol. 2013; 47(3):957–966. https://doi.org/10.1007/s12035-012-8392-4 PMID: 23315268

81. Jiang Q, Ho YY, Hao L, Berrios CN, Chakravarti A. Copy number variants in candidate genes are genetic modifiers of Hirschsprung disease. PLoS One. 2011; 6(6):e21219. https://doi.org/10.1371/journal.pone.0021219 PMID: 21712996

82. Tang W, Tang J, He J, Zhou Z, Qin Y, Qin J, et al. SLIT2/ROBO1-miR-218-1-RET/PLAG1: a new disease pathway involved in Hirschsprung's disease. Journal of cellular and molecular medicine. 2015; 19(6):1197–1207. https://doi.org/10.1111/jcmm.12454 PMID: 25786906

83. Clayton-Smith J, Laan L. Angelman syndrome: a review of the clinical and genetic aspects. Journal of Medical Genetics. 2003; 40(2):87–95. https://doi.org/10.1136/jmg.40.2.87 PMID: 12566516

84. Bird LM. Angelman syndrome: review of clinical and molecular aspects. Application of Clinical Genetics. 2014; 7. https://doi.org/10.2147/TACG.S57386 PMID: 24876791

85. Summers JA, Allison D, Lynch P, Sandier L. Behaviour problems in Angelman syndrome. Journal of Intellectual Disability Research. 1995; 39(2):97–106. https://doi.org/10.1111/j.1365-2788.1995.tb00477.x PMID: 7787388

86. Gilfillan GD, Selmer KK, Roxrud I, Smith R, Kyllerman M, Eiklid K, et al. SLC9A6 mutations cause X-linked mental retardation, microcephaly, epilepsy, and ataxia, a phenotype mimicking Angelman syndrome. The American Journal of Human Genetics. 2008; 82(4):1003–1010. https://doi.org/10.1016/j.ajhg.2008.01.013 PMID: 18342287

87. Pääbo S. The human condition—a molecular approach. Cell. 2014; 157(1):216–226. https://doi.org/10.1016/j.cell.2013.12.036 PMID: 24679537

88. Li M, Tian S, Yeung CK, Meng X, Tang Q, Niu L, et al. Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. Scientific reports. 2014; 4:4678. https://doi.org/10.1038/srep04678 PMID: 24728479

89. Shi Y, Xia YY, Wang L, Liu R, Khoo KS, Feng ZW. Neural cell adhesion molecule modulates mesenchymal stromal cell migration via activation of MAPK/ERK signaling. Experimental cell research. 2012; 318(17):2257–2267. https://doi.org/10.1016/j.yexcr.2012.05.029 PMID: 22683856

90. Azizan EA, Poulsen H, Tuluc P, Zhou J, Clausen MV, Lieb A, et al. Somatic mutations in ATP1A1 and CACNA1D underlie a common subtype of adrenal hypertension. Nature genetics. 2013; 45(9):1055–1060. https://doi.org/10.1038/ng.2716 PMID: 23913004

91. Scholl UI, Goh G, Stölting G, De Oliveira RC, Choi M, Overton JD, et al. Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. Nature genetics. 2013; 45(9):1050–1054. https://doi.org/10.1038/ng.2695 PMID: 23913001

92. Satheesh SV, Kunert K, Rüttiger L, Zuccotti A, Schönig K, Friauf E, et al. Retrocochlear function of the peripheral deafness gene Cacna1d. Human molecular genetics. 2012; p. dds217.

93. Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JB, Backström N, et al. Molecular evolution of genes in avian genomes. Genome biology. 2010; 11(6):R68. https://doi.org/10.1186/gb-2010-11-6-r68 PMID: 20573239

94. Begni S, Popoli M, Moraschi S, Bignotti S, Tura G, Gennarelli M. Association between the ionotropic glutamate receptor kainate 3 (GRIK3) ser310ala polymorphism and schizophrenia. Molecular psychiatry. 2002; 7(4):416–418. https://doi.org/10.1038/sj.mp.4000987 PMID: 11986986

95. Guo SZ, Huang K, Shi YY, Tang W, Zhou J, Feng GY, et al. A case-control association study between the GRID1 gene and schizophrenia in the Chinese Northern Han population. Schizophrenia research. 2007; 93(1):385–390. https://doi.org/10.1016/j.schres.2007.03.007 PMID: 17490860

96. Chen PL, Avramopoulos D, Lasseter VK, McGrath JA, Fallin MD, Liang KY, et al. Fine mapping on chromosome 10q22-q23 implicates Neuregulin 3 in schizophrenia. The American Journal of Human Genetics. 2009; 84(1):21–34. https://doi.org/10.1016/j.ajhg.2008.12.005 PMID: 19118813

97. Sahoo T, Theisen A, Rosenfeld JA, Lamb AN, Ravnan JB, Schultz RA, et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. Genetics in Medicine. 2011; 13(10):868–880. https://doi.org/10.1097/GIM.0b013e3182217a06 PMID: 21792059

98. Bozdagi O, Sakurai T, Papapetrou D, Wang X, Dickstein DL, Takahashi N, et al. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. Molecular autism. 2010; 1(1):1. https://doi.org/10.1186/2040-2392-1-15

99. Stewart SE, Yu D, Scharf JM, Neale BM, Fagerness JA, Mathews CA, et al. Genome-wide association study of obsessive-compulsive disorder. Molecular psychiatry. 2013; 18(7):788–798. https://doi.org/10.1038/mp.2012.85 PMID: 22889921

100. Won H, Lee HR, Gee HY, Mah W, Kim JI, Lee J, et al. Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function. Nature. 2012; 486(7402):261–265. https://doi.org/10.1038/nature11208 PMID: 22699620

101. Mattheisen M, Samuels JF, Wang Y, Greenberg BD, Fyer AJ, McCracken JT, et al. Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. Molecular psychiatry. 2015; 20(3):337–344. https://doi.org/10.1038/mp.2014.43 PMID: 24821223

102. Li Y, Wang GD, Wang MS, Irwin DM, Wu DD, Zhang YP. Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. Genome biology and evolution. 2014; 6(11):3115–3121. https://doi.org/10.1093/gbe/evu245 PMID: 25377939

103. Jarvis ED. Evolution of brain pathways for vocal learning in birds and humans. Birdsong, speech, and language. 2014; p. 63–108.

104. Wang R, Chen CC, Hara E, Rivas MV, Roulhac PL, Howard JT, et al. Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. Journal of Comparative Neurology. 2015; 523(6):892–906. https://doi.org/10.1002/cne.23719 PMID: 25424606

105. Mozzi A, Forni D, Clerici M, Pozzoli U, Mascheretti S, Guerini FR, et al. The evolutionary history of genes involved in spoken and written language: beyond FOXP2. Scientific reports. 2016; 6. https://doi.org/10.1038/srep22157 PMID: 26912479

106. Stein E, Tessier-Lavigne M. Hierarchical organization of guidance receptors: silencing of netrin attraction by slit through a Robo/DCC receptor complex. Science. 2001; 291(5510):1928–1938. https://doi.org/10.1126/science.1058445 PMID: 11239147

107. Dickinson RE, Duncan WC. The SLIT–ROBO pathway: a regulator of cell function with implications for the reproductive system. Reproduction. 2010; 139(4):697–704. https://doi.org/10.1530/REP-10-0017 PMID: 20100881

108. Manitt C, Mimee A, Eng C, Pokinko M, Stroh T, Cooper HM, et al. The netrin receptor DCC is required in the pubertal organization of mesocortical dopamine circuitry. The Journal of Neuroscience. 2011; 31(23):8381–8394. https://doi.org/10.1523/JNEUROSCI.0606-11.2011 PMID: 21653843

109. Grant A, Fathalli F, Rouleau G, Joober R, Flores C. Association between schizophrenia and genetic variation in DCC: a case–control study. Schizophrenia research. 2012; 137(1):26–31. https://doi.org/10.1016/j.schres.2012.02.023 PMID: 22418395

110. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. 2011; 471(7337):216–219. https://doi.org/10.1038/nature09774 PMID: 21390129

111. Leggere JC, Saito Y, Darnell RB, Tessier-Lavigne M, Junge HJ, Chen Z. NOVA regulate Dcc alternative splicing during neuronal migration and axon guidance in the spinal cord. eLife. 2016; 5:e14264. https://doi.org/10.7554/eLife.14264 PMID: 27223328

112. Antoine-Bertrand J, Duquette PM, Alchini R, Kennedy TE, Fournier AE, Lamarche-Vane N. p120RasGAP Protein Mediates Netrin-1 Protein-induced Cortical Axon Outgrowth and Guidance. Journal of Biological Chemistry. 2016; 291(9):4589–4602. https://doi.org/10.1074/jbc.M115.674846 PMID: 26710849

113. Balemans MC, Ansar M, Oudakker AR, van Caam AP, Bakker B, Vitters EL, et al. Reduced Euchromatin histone methyltransferase 1 causes developmental delay, hypotonia, and cranial abnormalities associated with increased bone gene expression in Kleefstra syndrome mice. Developmental biology. 2014; 386(2):395–407. https://doi.org/10.1016/j.ydbio.2013.12.016 PMID: 24362066

114. Willemsen MH, Vulto-van Silfhout AT, Nillesen WM, Wissink-Lindhout WM, van Bokhoven H, Philip N, et al. Update on Kleefstra syndrome. Molecular syndromology. 2012; 2(3-5):202–212. https://doi.org/10.1159/000335648 PMID: 22670141

115. Castellano S, Parra G, Sánchez-Quinto FA, Racimo F, Kuhlwilm M, Kircher M, et al. Patterns of coding variation in the complete exomes of three Neandertals. Proceedings of the National Academy of Sciences. 2014; 111(18):6666–6671. https://doi.org/10.1073/pnas.1405138111

116. Ozomaro U, Cai G, Kajiwara Y, Yoon S, Makarov V, Delorme R, et al. Characterization of SLITRK1 variation in obsessive-compulsive disorder. PloS one. 2013; 8(8):e70376. https://doi.org/10.1371/journal.pone.0070376 PMID: 23990902

117. Alexander J, Potamianou H, Xing J, Deng L, Karagiannidis I, Tsetsos F, et al. Targeted re-sequencing approach of candidate genes implicates rare potentially functional variants in Tourette Syndrome etiology. Frontiers in Neuroscience. 2016; 10. https://doi.org/10.3389/fnins.2016.00428

118. Kmoch S, Hartmannová H, Stibùrková B, Krijt J, Zikánová M, Šebesta I. Human adenylosuccinate lyase (ADSL), cloning and characterization of full-length cDNA and its isoform, gene structure and molecular basis for ADSL deficiency in six patients. Human molecular genetics. 2000; 9(10):1501–1513. https://doi.org/10.1093/hmg/9.10.1501 PMID: 10888601

119. Jurecka A, Zikanova M, Kmoch S, Tylki-Szymańska A. Adenylosuccinate lyase deficiency. Journal of inherited metabolic disease. 2015; 38(2):231–242. https://doi.org/10.1007/s10545-014-9755-y PMID: 25112391

120. Vigne JD. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. Comptes rendus biologies. 2011; 334(3):171–181. https://doi.org/10.1016/j.crvi.2010.12.009 PMID: 21377611

121. Albert FW, Somel M, Carneiro M, Aximu-Petri A, Halbwax M, Thalmann O, et al. A comparison of brain gene expression levels in domesticated and wild animals. PLoS Genet. 2012; 8(9):e1002962. https://doi.org/10.1371/journal.pgen.1002962 PMID: 23028369

122. Pendleton AL, Shen F, Taravella AM, Emery S, Veeramah KR, Boyko AR, et al. Selective sweep analysis using village dogs highlights the pivotal role of the neural crest in dog domestication. bioRxiv. 2017; p. 118794.

123. Stanley CE, Kulathinal RJ. Genomic signatures of domestication on neurogenetic genes in Drosophila melanogaster. BMC evolutionary biology. 2016; 16(1):1. https://doi.org/10.1186/s12862-015-0580-1

124. Wang Gd, Zhai W, Yang Hc, Fan Rx, Cao X, Zhong L, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. Nature communications. 2013; 4:1860. https://doi.org/10.1038/ncomms2814 PMID: 23673645

125. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 2014; 345 (6200):1074–1079. https://doi.org/10.1126/science.1253714 PMID: 25170157

126. Hare B, Woods V. The genius of dogs: how dogs are smarter than you think. Penguin; 2013.

127. Okanoya K. Behavioural factors governing song complexity in Bengalese finches. International Journal of Comparative Psychology. 2012; 25(1).

128. Shuldiner E, Koch IJ, Kartzinel RY, Hogan A, Brubaker L, Wanser S, et al. Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. Science Advances. 2017; 3(7):e1700398. https://doi.org/10.1126/sciadv.1700398 PMID: 28776031

129. Cruz F, Vilà C, Webster MT. The legacy of domestication: accumulation of deleterious mutations in the dog genome. Molecular biology and evolution. 2008; 25(11):2331–2336. https://doi.org/10.1093/molbev/msn177 PMID: 18689870

130. Björnerfeldt S, Webster MT, Vilà C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Research. 2006; 16(8):990–994. https://doi.org/10.1101/gr.5117706 PMID: 16809672

131. Wang Z, Yonezawa T, Liu B, Ma T, Shen X, Su J, et al. Domestication relaxed selective constraints on the yak mitochondrial genome. Molecular biology and evolution. 2011; 28(5):1553–1556. https://doi.org/10.1093/molbev/msq336 PMID: 21156878

132. Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends in Genetics. 2006; 22(3):126–131. https://doi.org/10.1016/j.tig.2006.01.004 PMID: 16443304

133. Wang R. Dissecting the Genetic Basis of Convergent Complex Traits Based on Molecular Homoplasy. Duke University. test; 2011.

134. Webb DM, Zhang J. FoxP2 in song-learning birds and vocal-learning mammals. Journal of Heredity. 2005; 96(3):212–216. https://doi.org/10.1093/jhered/esi025 PMID: 15618302

135. Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, et al. The bonobo genome compared with the chimpanzee and human genomes. Nature. 2012; 486(7404):527–531. https://doi.org/10.1038/nature11128 PMID: 22722832

136. Heyne HO, Lautenschläger S, Nelson R, Besnier F, Rotival M, Cagan A, et al. Genetic influences on brain gene expression in rats selected for tameness and aggression. Genetics. 2014; 198(3):1277–1290. https://doi.org/10.1534/genetics.114.168948 PMID: 25189874

137. Cagan A, Theunert C, Laayouni H, Santpere G, Pybus M, Casals F, et al. Natural Selection in the Great Apes. Molecular Biology and Evolution. 2016; 33(12):3268–3283. https://doi.org/10.1093/molbev/msw215 PMID: 27795229

138. R Core Team. R: A Language and Environment for Statistical Computing; 2013. Available from: http://www.R-project.org/.

139. Kalinka AT. The probability of drawing intersections: extending the hypergeometric distribution. arXiv preprint arXiv:13050717. 2013;.

140. Pilot M, Malewski T, Moura AE, Grzybowski T, Oleński K, Kamiński S, et al. Diversifying selection between pure-breed and free-breeding dogs inferred from genome-wide SNP analysis. G3: Genes| Genomes| Genetics. 2016; p. g3–116.

Chapter 2

frontiers in Neuroscience

PERSPECTIVE
published: 14 June 2016
doi: 10.3389/fnins.2016.00271

Implications of Oxytocin in Human Linguistic Cognition: From Genome to Phenome

Constantina Theofanopoulou *

Department of General Linguistics, University of Barcelona, Barcelona, Spain

The neurohormone oxytocin (OXT) has been found to mediate the regulation of complex socioemotional cognition in multiple ways both in humans and other animals. Recent studies have investigated the effects of OXT in different levels of analysis (from genetic to behavioral) chiefly targeting its impact on the social component and only indirectly indicating its implications in other components of our socio-interactive abilities. This article aims at shedding light onto how OXT might be modulating the multimodality that characterizes our higher-order linguistic abilities (vocal-auditory-attentional-memory-social systems). Based on evidence coming from genetic, EEG, fMRI, and behavioral studies, I attempt to establish the promises of this perspective with the goal of stressing the need for neuropeptide treatments to enter clinical practice.

Keywords: oxytocin, language, alpha rhythm, OXTR, POU3F2, LNPEP, FOXP2, CNTNAP2

OPEN ACCESS

Edited by:
Mitsuhiro Kawata, Kyoto Prefectural University of Medicine, Japan

Reviewed by:
Yoichi Ueta, University of Occupational and Environmental Health, Japan
Haruhiro Higashida, Kanazawa University Research Center for Child Mental Development, Japan

***Correspondence:**
Constantina Theofanopoulou
constantinaki@hotmail.com

Specialty section:
This article was submitted to Neuroendocrine Science, a section of the journal Frontiers in Neuroscience

Received: *06 March 2016*
Accepted: *31 May 2016*
Published: *14 June 2016*

Citation:
Theofanopoulou C (2016) Implications of Oxytocin in Human Linguistic Cognition: From Genome to Phenome. Front. Neurosci. 10:271. doi: 10.3389/fnins.2016.00271

INTRODUCTION

The nine amino acid peptide oxytocin (OXT) is involved in an array of physiological and pathophysiological processes, with some of those most commonly reported in the literature being pregnancy and uterine contractions, milk ejection, sexual activity, pain modulation, social interaction and bonding, parental care, and attention to socially-relevant stimuli (for a good review see Meyer-Lindenberg et al., 2011). From another perspective, malfunctions of the oxytocinergic system have been reported in cases of Autism Spectrum Disorder, Schizophrenia, Obsessive Compulsive Disorder, Phobia, Prader-Willi Syndrome and Williams Syndrome, providing strong functional links to the social and emotional modules that all these cases share (Leckman et al., 1994; Lopatina et al., 2012; De Berardis et al., 2013; Grinevich et al., 2015; Haas and Smith, 2015). This broad perspective of the literature indicates that OXT impacts a wide spectrum of neurobehavioral systems.

Here I put forth the hypothesis that OXT also has a significant role in our linguistic abilities, ranging from modulating genes involved in spoken-language acquisition to modulating our motivation to communicate. In building this hypothesis, I follow an approach I have argued in Theofanopoulou and Boeckx (2015) in the context of cognitive phylogenies, where for a hypothesis to be valid in the Language Sciences, there needs to be evidence at multiple levels of biological organization, from genetics to ultimately the behavioral level (Fisher, 2015). Thus, I appeal to relevant findings from a multitude of studies, touching upon all the following levels of analysis: genome, connectome, dynome (brain oscillations), cognome, and phenome (See **Figure 1**). I also develop my hypothesis from a translational viewpoint among non-human animal studies and humans, including the molecular studies of OXT to its social functions in communication. I conclude that OXT most

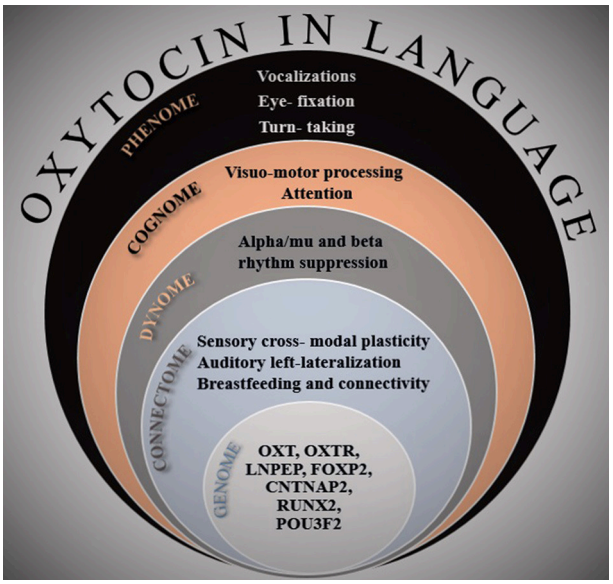


FIGURE 1 | A multi-dimensional illustration of the evidence presented in the paper. At every level of analysis, the most important findings that are related to the role of oxytocin in linguistic cognition are listed.

probably globally affects brain components that are tightly interwoven with the pinnacle of our social expressions, namely the sensory, motor, and more cognitive facets of our linguistic abilities (auditory, vocal, attention, and memory systems).

GENOME: OXT MODULATES GENES INVOLVED IN SPOKEN-LANGUAGE ACQUISITION

Apart from the aforementioned actions of OXT, what is of most relevance for the present article is its key role in several developmental processes that subserve the acquisition of our higher cognitive skills. Oxytocin-mediated, experience-dependent cross-modal plasticity in the sensory cortices during early development (Zheng et al., 2014) and the left-lateralized expression of OXT in the auditory cortex of the mouse brain (Marlin et al., 2015) suggest that OXT pathways are highly pertinent to understanding the sensory ontogeny of our linguistic communication. For humans, epigenetic misregulation of the OXTR via aberrant gene silencing with DNA methylation has implicated OXT in the development of Autism Spectrum Disorder, where deficits in language performance are included in its core phenotype (Gregory et al., 2009). A potential mechanism is that epigenetic DNA methylation of the oxytocin receptor gene (OXTR) is associated with neural activity and functional coupling of neurons (Puglia et al., 2015). Thus, the aberrant OXTR expression by methylation could be impacting neural activity and neuronal coupling in language performance.

An even more possible direct genetic link between OXT and our linguistic capacities is evidenced in the robust findings with genes known to be necessary for normal language development, namely in the FOXP2-CNTNAP2 pathway. To begin with, interaction between OXT and CNTNAP2 in critical developmental windows has been shown in a mouse model of autism (Peñagarikano et al., 2015). FOXP2 regulates CNTNAP2 expression, and CNTNAP2 has been linked to complex neurological disorders, including language impairment, autism, dyslexia, schizophrenia, and depression, with Single Nucleotide Polymorphisms (SNPs) having been associated with specific language endophenotypes (see Rodenas-Cuadrado et al., 2014 for review).

Another link between OXT and FOXP2 is provided through LNPEP, the peptidase that metabolizes oxytocin, located on chromosome 5q15 (for more details on LNPEP see Ebstein et al., 2012). Vernes et al. (2007) identified genomic sites directly bound by FOXP2 protein in native chromatin of human neuron-like cells, and LNPEP was among the genes with the most robust and consistent binding. LNEP functionally regulates synaptic transmission and formation.

A third potential interaction between OXT and FOXP2 may occur by two other genes related to language: (i) RUNX2 and (ii) POU3F2 (Benítez-Burraco and Boeckx, 2014, 2015). For RUNX2, –a critical transcription factor for osteoblast formation-, Tamma et al. (2009) found that it was differentially regulated in OXT knockout mice. RUNX2 is connected to many genes that are essential not only for brain and language development, but also for bone formation (Boeckx and Benítez-Burraco, 2015). A direct interaction between RUNX2 and FOXP2 has been experimentally demonstrated in the context of endochondral ossification (Zhao et al., 2015), a finding further reinforced by Gascoyne et al. (2015), who added FOXP2 to the list of established osteoblast and chondrocyte transcription factors (such as RUNX2). Significantly, the action of OXT on osteoblast maturation (Di Benedetto et al., 2014) and its implication in an osteogenic network that supports the development of our language-ready brain (and skull) may provide genetic evidence for the hypothesis that OXT may directly foster encephalization and our craniofacial phenotype (Carter, 2014). Last but not least, both OXT and RUNX2 have been found to be strongly connected to the Vitamin D endocrine system (Prüfer and Jirikowski, 1997; Han et al., 2013; Patrick and Ames, 2014), which has been proposed to explain the genetics and epidemiology of Autism (Cannell, 2008).

Concerning POU3F2, a transcription factor, neuronal and endocrine components (including OXT) of the hypothalamic-pituitary axis have been shown to be critically dependent on POU3F2 action (Nakai et al., 1995; Schonemann et al., 1995; Burbach et al., 2001). POU3F2 also regulates FOXP2 gene expression in a human-specific manner (Maricic et al., 2013). Crucially, the fact that in all three genes, OXTR, POU3F2, and FOXP2, there have been identified signs of positive selection in human or recent hominin evolution (Enard et al., 2002; Maricic et al., 2013; Schaschl et al., 2015), reinforces the idea that these evolutionary changes might be partially responsible for the emergence of aspects of our species-specific cognitive and linguistic abilities.

CONNECTOME

Recent studies have implicated OXT in brain development and plasticity. Specifically, the oxytocinergic brain system has been described to undergo major morphological alterations that modify the conformation of its neurons and glia and its synaptic inputs in a stimulus-dependent manner (Theodosis, 2002). The bulk of the evidence coming from studies in mice, rats and praire voles elucidates the significant role OXT plays in shaping different pathways of the brain (see Carter, 2003 for review). Importantly, the expression of the OXTR displays a particular maturational progression in the brain of the developing rat that could be classified in two types: transient expression during early postnatal development and constant abundant expression mediating neuronal transmission in the mature brain (Yoshimura et al., 1996). Similarly in mice, neocortical OXTR binding exhibits a transient peak in early postnatal periods, when extensive synaptic proliferation and pruning takes place (Hammock and Levitt, 2013).

These findings along with the ones that address the effect of the maturation of the OXT system on sensory—and not only socio-sexual- aspects could exemplify why early postnatal life is indeed a sensitive period for OXT in modeling circuits that are eventually responsible for sensory performance. Additional insight can be gained from comparative data on mice: Zheng et al. (2014) found that OXT promotes excitatory synaptic transmission in the sensory cortices at a much earlier stage than the hitherto understood functions of OXT in social and emotional contexts and, notably, Marlin et al. (2015) found that both OXT receptors and projections from hypothalamic OXT-producing neurons are present in the auditory cortex of mice, with the former being more numerous on the left side than on the right, something that could be telling for lateralization in human language development (Theofanopoulou, 2015 and references therein).

In humans, it has not yet been experimentally established how early adjustments of the OXT system influence the neuronal and synaptic substrates that underlie the sensory and cognitive modules of our language-ready brain. The only (rough) conclusions we can deduce from the literature are based on comparisons between infants that have or have not been breastfed and concomitant brain changes. On the grounds that OXT is stable in milk and that OXT in maternal blood can be transferred to milk and then to neonates (Takeda et al., 1986), we would expect that lactation goes hand-in-hand with proliferating brain connectivity. At least, some evidence suggests so: Deoni et al. (2013) showed an association between early exclusive breastfeeding with increased development in late maturing white matter regions (interestingly also near BA44, traditionally linked to language). Tellingly, breastfed children also showed improved receptive language scores compared to formula-fed children. Moreover, Khedr et al. (2004) found that visual evoked potential (FVEP), brainstem auditory evoked potential (BAEP), and somatosensory evoked potential (SSEP) are more mature in breastfed infants relative to formula-fed infants at 1-year of age, something suggestive of the importance of breastfeeding in early development. I propose that an important molecule and factor could be the high concentration in OXT in breast milk and also its release during skin-to-skin contact over breastfeeding (Uvnäs-Moberg et al., 2015). Furthermore, the aforementioned results (see also Kafouri et al., 2013 and Isaacs et al., 2010) mesh well with recent studies showing that autism is to a great extent correlated with inefficient breastfeeding, by cause of lack of interest in milk-suckling (Williams et al., 2000; Gallup and Hobbs, 2011; Al-Farsi et al., 2012; Steinman and Mankuta, 2013). A deeper understanding of the complex OXT feedback loop between mother and infant in breastfeeding could be reached if we additionally take into account that the perturbation of the system might be actually stemming from the mother. Indeed, birth complications (Brimacombe et al., 2007) due to low OXT levels and stressful-depressive mother care have long been associated with autism (see Uvnäs-Moberg et al., 2015 for an excellent review on the short- and long-term effects of breastfeeding and skin-to-skin contact between mother and infant, explained via OXT release). According to this thread of interpretation, traditional psychological theories on the role of the "refrigerator-mother" in the etiology of autism could now be construed on a neuroendocrine basis.

Another important issue at the level of the connectome is the loci where OXT is expressed in the brain. In humans, OXT is dispersed from the magnocellular neurons in the paraventricular and supraoptic nuclei of the hypothalamus to practically throughout the brain: including the amygdala, the hippocampus, the striatum, the brainstem, the cerebellum, the insula, the suprachiasmatic nucleus, the septum, the bed nucleus of stria terminalis, the globus pallidus, the substantia nigra pars compacta, the ventral tegmental are, the spinal cord, and to neocortical areas traditionally associated with "language," such as the prefrontal cortex, the anterior cingulate cortex and the precuneus (Lee et al., 2009, 2010; Ma et al., 2016). Even though it is important to find out "where" OXT is expressed in the brain, a mere locationist approach cannot enlighten our understanding of "how" OXT gives rise to cognitive sub-processes mechanistically (Theofanopoulou and Boeckx, 2015). At the following level of analysis (i.e., the dynome) the direct effects of OXT administration on brain rhythms and how this translates into specific cognitive processes (i.e., the cognome) will be illustrated.

DYNOME—COGNOME

Only very recently attempts have been made to link the action of OXT with a rhythmic correlate in the human brain that would make some sense in terms of its cognitive significance. In early experimental attempts of pure behavioral paradigms (e.g., "trust" experiments, for example: Baumgartner et al., 2008), OXT was not implicated at a granularity level that could be matched with the (de)activation of a specific oscillatory band. It was not until 2009, when Kéri and Benedek examined the effect of OXT on the perception of biological vs. non-biological motion stimuli, that a venue for associating OXT modulation to neural activity was opened (Kéri and Benedek, 2009). Specifically, Kéri and Benedek found that OXT enhances the ability to detect biological motion in noise, whereas no such effect turned up when detecting a

rotating shape. This led Perry et al. (2010) to tentatively link these results with the alpha/mu and beta brain rhythms, which have been shown to be suppressed while observing actions executed by someone else (Muthukumaraswamy and Johnson, 2004; Lange et al., 2015). Characteristically, alpha/mu and beta rhythms have been found to be desynchronized reinforcing the efficiency of the mirror neuron system, which in humans is activated not only when observing biological actions, but also at all levels of communicative interactions (see Pineda, 2005 for a review). This is more than pertinent to the scope of this article, since for linguistic communication interplay to happen, it is necessary not only to perceive biological movements (lip-movements, tongue-movements, formant transitions, hand gestures, and eye movements), but to couple them with the auditory input and whence make out the multidimensional meaning of the compound "linguistic" input. As I have put forward elsewhere (Theofanopoulou, 2015), this interplay should be mediated by an attentional mechanism that keeps track of all these distinct rhythmic stimuli. It should not take us by surprise then that an overall decrease of the aforesaid rhythms has also been linked to increased demands of attention and memory (Klimesch, 2012). Importantly, after OXT administration, alpha/mu and beta rhythms had a general suppressive effect that was widespread across the scalp (viz not only on brain areas of the somato-motor cortex), something that was interpreted as an effect on a broader network, in which mirror/motor and attentional mechanisms can be with difficultly disentangled (Perry et al., 2010). A similar experiment was conducted by Singh et al. (2015), also in Schizophrenia patients, and replicated the diffused effect of OXT in the brain. Lastly, Hepker (2016) tested how OXT affects mirror neuron activity in a hand-gesture experiment and encountered greater mu rhythm suppression, in accordance with other experiments, but this time for a biological movement directly involved in language processing.

To the best of my knowledge, there are no studies yet showing that OXT has a direct effect on the rhythmic patterns in a purely linguistic task. But as put forth in Theofanopoulou (2016) there are several reasons to expect so. Firstly, alpha/mu and beta band suppression have been shown to coordinate the rhythms partaking not only in motor but also in auditory (speech) (Obleser and Weisz, 2012) processing and OXT seems to support this multimodality, considering that it has been found to increase not only in response to biological motions, but also to vocalizations alone (Seltzer et al., 2010) and to attenuate the human acoustic startle response (Ellenbogen et al., 2014). Secondly, in autism alpha-band deployment was shown to be severely impaired, giving rise to increased distraction (Oberman et al., 2005, 2008; Murphy et al., 2014; see also Moran and Hong, 2011, for similar findings in schizophrenia). Here magnetoencephalography (MEG) studies showing atypical auditory responses in patients with autism are also of relevance: for example, in autistic patients stronger responses to nonspeech than speech sounds (Yau et al., 2015), delayed (Roberts et al., 2010), and atypically lateralized (Orekhova et al., 2012) neuromagnetic auditory field responses compared to controls were observed. These experiments in conjunction with the irregularities observed in the oxytocinergic system in autism make it plausible that OXT might in part modulate the brain rhythms in language-processing.

PHENOME

In behavioral experiments OXT has been engaged in a surfeit of different complex tasks that can be difficult to decompose for the aims of this article. Accordingly, only experiments that are informative for different facets of linguistic processing will be mentioned.

OXT has been loosely associated with "communicative" functions (Yamasue, 2013) that only recently have been broken down into processes that correspond to more specific linguistic processes. For instance, Seltzer et al. (2010) found that children under stress show increased OXT levels after hearing maternal vocalizations and Watanabe et al. (2014) showed that intranasal OXT administration to autism-patients affects their decisions about social information with conflicting verbal and non-verbal contents. Lastly, Ellenbogen et al. (2014) found that intranasal oxytocin attenuates the human acoustic startle response independent of emotional modulation.

However, most data come from studies involving OXT in eye-gaze enhancement suggesting its plausible role in interpersonal communication (Guastella et al., 2008) and in inferring the mental state of others (Domes et al., 2007). Gamer (2010) explains that OXT increased the proportion of fixation changes toward the eyes across all expressions, and did not directly affect the efficiency of processing emotional faces *per se*. In light of studies clarifying the importance of eye gaze in the modulation of speech and co-speech gesture (Holler et al., 2014, 2015), we can better appraise why in most cases the communicative deficits in autism derive from an abnormal fixation to the mouth region of the interlocutor, instead of the eye region (Pelphrey et al., 2005; Neumann et al., 2006). Tellingly, for therapeutic concerns, Andari et al. (2010) found that OXT selectively increased autism patients' gazing time on the eye region, improving their social performance.

In a similar vein, Ebitz and Platt (2014) further argue that these emitted eye-signals, regulated by OXT, provoke OXT-release back in the receiver, increasing eye contact and proximity seeking, establishing in this way a back-and-forward loop that strongly underlies communicative functions. This cascade of reciprocal OXT-secretion might, in other words, give a neurohormonal basis to the "turn-taking" roots of our linguistic capacity, recently highlighted from an evolutionary perspective (Levinson, 2016).

CONCLUSION

In this article I attempted to draw attention to the potential implications of the neurohormone OXT in the context of language. Even though its role in purely linguistic matters has so far been overlooked, there is already a plethora of evidence strongly suggesting that a better understanding of its function could be rewarding. Results from experiments at different levels of analysis (from genetic to oscillatory and behavioral) suggest

that OXT could fit well in the recently addressed hypotheses that underline the "reward-learning" foundations of our linguistic capacities (see Berra, 2015 for a good review). However, till now only dopamine has been tested in linguistic tasks in humans (Ripollés et al., 2014) and widely in vocal-learning in zebra finches (reviewed in Simonyan et al., 2012). More genetic experiments on the effect of OXT on mice vocalizations and birdsongs in different paradigms (courtship, affiliative, fear, dam-puppies) and EEG studies on its impact on alpha/mu/beta rhythm suppression in a speech perception task would help to appreciate more the role of OXT in our high cognition and its possible therapeutic implications.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

Preparation of this work was supported by funds from the Spanish Ministry of Economy and Competitiveness (grant FFI2013-43823-P), as well as funds from the Generalitat de Catalunya (FI-grant-2016-2019).

ACKNOWLEDGMENTS

I am indebted to my advisor, Cedric Boeckx, for his constant support and help. I would also like to thank my co-advisor Erich Jarvis for his valuable comments and instructive discussions, Olga Peñagarikano for helpful information on the genetics of oxytocin and Sonja Vernes for very fruitful conversations on the genetics of language. Lastly, I am grateful to David Poeppel for his comments on how the rhythmic effect of oxytocin should be tested.

REFERENCES

Al-Farsi, Y. M., Al-Sharbati, M. M., Waly, M. I., Al-Farsi, O. A., Al-Shafaee, M. A., Al-Khaduri, M. M., et al. (2012). Effect of suboptimal breast-feeding on occurrence of autism: a case-control study. *Nutrition* 28, e27–e32. doi: 10.1016/j.nut.2012.01.007

Andari, E., Duhamel, J.-R., Zalla, T., Herbrecht, E., Leboyer, M., and Sirigu, A. (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4389–4394. doi: 10.1073/pnas.0910249107

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009

Benítez-Burraco, A., and Boeckx, C. (2014). FOXP2, retinoic acid, and language: a promising direction. *Front. Cell. Neurosci.* 8:387. doi: 10.3389/fncel.2014.00387

Benítez-Burraco, A., and Boeckx, C. (2015). Possible functional links among brain- and skull-related genes selected in modern humans. *Front. Psychol.* 6:794. doi: 10.3389/fpsyg.2015.00794

Berra, I. (2015). Emotions in vocal learning. The continuity behind the convergence. *Reti Saperi Linguaggi.* 1, 167–184. doi: 10.12832/81296

Boeckx, C., and Benítez-Burraco, A. (2015). Osteogenesis and neurogenesis: a robust link also for language evolution. *Front. Cell. Neurosci.* 9:291. doi: 10.3389/fncel.2015.00291

Brimacombe, M., Ming, X., and Lamendola, M. (2007). Prenatal and birth complications in autism. *Matern. Child Health J.* 11, 73–79. doi: 10.1007/s10995-006-0142-7

Burbach, J. P. H., Luckman, S. M., Murphy, D., and Gainer, H. (2001). Gene regulation in the magnocellular hypothalamo-neurohypophysial system. *Physiol. Rev.* 81, 1197–1267.

Cannell, J. J. (2008). Autism and vitamin D. *Med. Hypotheses* 70, 750–759. doi: 10.1016/j.mehy.2007.08.016

Carter, C. S. (2003). Developmental consequences of oxytocin. *Physiol. Behav.* 79, 383–397.

Carter, C. S. (2014). Oxytocin pathways and the evolution of human behavior. *Annu. Rev. Psychol.* 65, 17–39. doi: 10.1146/annurev-psych-010213-115110

De Berardis, D., Marini, S., Iasevoli, F., Tomasetti, C., de Bartolomeis, A., Mazza, M., et al. (2013). The role of intranasal oxytocin in the treatment of patients with schizophrenia: a systematic review. *CNS Neurol. Disord. Drug Targets* 12, 252–264.

Deoni, S. C. L., Dean, D. C., Piryatinsky, I., O'Muircheartaigh, J., Waskiewicz, N., Lehman, K., et al. (2013). Breastfeeding and early white matter development: a cross-sectional study. *Neuroimage* 82, 77–86. doi: 10.1016/j.neuroimage.2013.05.090

Di Benedetto, A., Sun, L., Zambonin, C. G., Tamma, R., Nico, B., Calvano, C. D., et al. (2014). Osteoblast regulation via ligand-activated nuclear trafficking of the oxytocin receptor. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16502–16507. doi: 10.1073/pnas.1419349111

Domes, G., Heinrichs, M., Michel, A., Berger, C., and Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biol. Psychiatry* 61, 731–733. doi: 10.1016/j.biopsych.2006.07.015

Ebitz, R. B., and Platt, M. M. (2014). An evolutionary perspective on the behavioral consequences of exogenous oxytocin application. *Front. Behav. Neurosci.* 7:225. doi: 10.3389/fnbeh.2013.00225

Ebstein, R. P., Knafo, A., Mankuta, D., Chew, S. H., and Lai, P. S. (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. *Horm. Behav.* 61, 359–379. doi: 10.1016/j.yhbeh.2011.12.014

Ellenbogen, M. A., Linnen, A.-M., Cardoso, C., and Joober, R. (2014). Intranasal oxytocin attenuates the human acoustic startle response independent of emotional modulation. *Psychophysiology* 51, 1169–1177. doi: 10.1111/psyp.12263

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418, 869–872. doi: 10.1038/nature01025

Fisher, S. E. (2015). "Translating the genome in human neuroscience," in *The Future of the Brain: Essays by the World's Leading Neuroscientists,* eds G. Marcus and J. Freeman (Princeton, NJ: Princeton University Press), 149–159.

Gallup, G. G., and Hobbs, D. R. (2011). Evolutionary medicine: bottle feeding, birth spacing, and autism. *Med. Hypotheses* 77, 345–346. doi: 10.1016/j.mehy.2011.05.010

Gamer, M. (2010). Does the amygdala mediate oxytocin effects on socially reinforced learning? *J. Neurosci.* 30, 9347–9348. doi: 10.1523/JNEUROSCI.2847-10.2010

Gascoyne, D. M., Spearman, H., Lyne, L., Puliyadi, R., Perez-Alcantara, M., Coulton, L., et al. (2015). The forkhead transcription factor FOXP2 is required for regulation of p21WAF1/CIP1 in 143B osteosarcoma cell growth arrest. *PLoS ONE* 10:e0128513. doi: 10.1371/journal.pone.0128513

Gregory, S. G., Connelly, J. J., Towers, A. J., Johnson, J., Biscocho, D., Markunas, C. A., et al. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Med.* 7:62. doi: 10.1186/1741-7015-7-62

Grinevich, V., Desarménien, M. G., Chini, B., Tauber, M., and Muscatelli, F. (2015). Ontogenesis of oxytocin pathways in the mammalian brain: late maturation and psychosocial disorders. *Front. Neuroanat.* 8:164. doi: 10.3389/fnana.2014.00164

Guastella, A. J., Mitchell, P. B., and Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biol. Psychiatry* 63, 3–5. doi: 10.1016/j.biopsych.2007.06.026

Haas, B. W., and Smith, A. K. (2015). Oxytocin, vasopressin, and Williams syndrome: epigenetic effects on abnormal social behavior. *Front. Genet.* 6:28. doi: 10.3389/fgene.2015.00028

Hammock, E. A. D., and Levitt, P. (2013). Oxytocin receptor ligand binding in embryonic tissue and postnatal brain development of the C57BL/6J mouse. *Front. Behav. Neurosci.* 7:195. doi: 10.3389/fnbeh.2013.00195

Han, M.-S., Che, X., Cho, G., Park, H.-R., Lim, K.-E., Park, N.-R., et al. (2013). Functional cooperation between vitamin D receptor and Runx2 in vitamin D-induced vascular calcification. *PLoS ONE* 8:e83584. doi: 10.1371/journal.pone.0083584

Hepker, M. (2016). "Effect of oxytocin administration on mirror neuron activation," in *Summer Research.* Paper 176. Available online at: http://soundideas.pugetsound.edu/summer_research/176

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S. D., and Özyürek, A. (2015). Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Soc. Cogn. Affect. Neurosci.* 10, 255–261. doi: 10.1093/scan/nsu047

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., and Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition* 133, 692–697. doi: 10.1016/j.cognition.2014.08.008

Isaacs, E. B., Fischl, B. R., Quinn, B. T., Chong, W. K., Gadian, D. G., and Lucas, A. (2010). Impact of breast milk on IQ, brain size and white matter development. *Pediatr. Res.* 67, 357–362. doi: 10.1203/PDR.0b013e3181d026da

Kafouri, S., Kramer, M., Leonard, G., Perron, M., Pike, B., Richer, L., et al. (2013). Breastfeeding and brain structure in adolescence. *Int. J. Epidemiol.* 42, 150–159. doi: 10.1093/ije/dys172

Kéri, S., and Benedek, G. (2009). Oxytocin enhances the perception of biological motion in humans. *Cogni. Affect. Behav. Neurosci.* 9, 237–241. doi: 10.3758/CABN.9.3.237

Khedr, E., Farghaly, W., Amry, S. E.-D., and Osman, A. (2004). Neural maturation of breastfed and formula-fed infants. *Acta Paediatr.* 93, 734–738. doi: 10.1111/j.1651-2227.2004.tb03011.x

Klimesch, W. (2012). α-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Lange, J., Pavlidou, A., and Schnitzler, A. (2015). Lateralized modulation of beta-band power in sensorimotor areas during action observation. *Front. Integr. Neurosci.* 9:43. doi: 10.3389/fnint.2015.00043

Leckman, J. F., Goodman, W. K., North, W. G., Chappell, P. B., Price, L. H., Pauls, D. L., et al. (1994). The role of central oxytocin in obsessive compulsive disorder and related normal behavior. *Psychoneuroendocrinology* 19, 723–749.

Lee, H.-J., Macbeth, A. H., Pagani, J. H., and Young, W. S. (2009). Oxytocin: the great facilitator of life. *Prog. Neurobiol.* 88, 127–151. doi: 10.1016/j.pneurobio.2009.04.001

Lee, H.-J., Pagani, J., and Young, W. S. (2010). Using transgenic mouse models to study oxytocin's role in the facilitation of species propagation. *Brain Res.* 1364, 216–224. doi: 10.1016/j.brainres.2010.08.042

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Lopatina, O., Inzhutova, A., Salmina, A. B., and Higashida, H. (2012). The roles of oxytocin and CD38 in social or parental behaviors. *Front. Neurosci.* 6:182. doi: 10.3389/fnins.2012.00182

Ma, Y., Shamay-Tsoory, S., Han, S., and Zink, C. F. (2016). Oxytocin and social adaptation: insights from neuroimaging studies of healthy and clinical populations. *Trends Cogn. Sci.* 20, 133–145. doi: 10.1016/j.tics.2015.10.009

Maricic, T., Günther, V., Georgiev, O., Gehre, S., Curlin, M., Schreiweis, C., et al. (2013). A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Mol. Biol. Evol.* 30, 844–852. doi: 10.1093/molbev/mss271

Marlin, B. J., Mitre, M., D'amour, J. A., Chao, M. V., and Froemke, R. C. (2015). Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* 520, 499–504. doi: 10.1038/nature14402

Meyer-Lindenberg, A., Domes, G., Kirsch, P., and Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat. Rev. Neurosci.* 12, 524–538. doi: 10.1038/nrn3044

Moran, L. V., and Hong, L. E. (2011). High vs low frequency neural oscillations in schizophrenia. *Schizophr. Bull.* 37, 659–663. doi: 10.1093/schbul/sbr056

Murphy, J. W., Foxe, J. J., Peters, J. B., and Molholm, S. (2014). Susceptibility to distraction in autism spectrum disorder: probing the integrity of oscillatory alpha-band suppression mechanisms. *Autism Res.* 7, 442–458. doi: 10.1002/aur.1374

Muthukumaraswamy, S. D., and Johnson, B. W. (2004). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology* 41, 152–156. doi: 10.1046/j.1469-8986.2003.00129.x

Nakai, S., Kawano, H., Yudate, T., Nishi, M., Kuno, J., Nagata, A., et al. (1995). The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse. *Genes Dev.* 9, 3109–3121.

Neumann, D., Spezio, M. L., Piven, J., and Adolphs, R. (2006). Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Soc. Cogn. Affect. Neurosci.* 1, 194–202. doi: 10.1093/scan/nsl030

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res. Cogn. Brain Res.* 24, 190–198. doi: 10.1016/j.cogbrainres.2005.01.014

Oberman, L. M., Ramachandran, V. S., and Pineda, J. A. (2008). Modulation of mu suppression in children with autism spectrum disorders in response to familiar or unfamiliar stimuli: the mirror neuron hypothesis. *Neuropsychologia* 46, 1558–1565. doi: 10.1016/j.neuropsychologia.2008.01.010

Obleser, J., and Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details.*Cereb. Cortex* 22, 2466–2477. doi: 10.1093/cercor/bhr325

Orekhova, E. V., Tsetlin, M. M., Butorina, A. V., Novikova, S. I., Gratchev, V. V., Sokolov, P. A., et al. (2012). Auditory cortex responses to clicks and sensory modulation difficulties in children with Autism Spectrum Disorders (ASD). *PLoS ONE* 7:e39906. doi: 10.1371/journal.pone.0039906

Patrick, R. P., and Ames, B. N. (2014). Vitamin D hormone regulates serotonin synthesis. Part 1: relevance for autism. *FASEB J.* 28, 2398–2413. doi: 10.1096/fj.13-246546

Pelphrey, K. A., Morris, J. P., and McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain* 128(Pt 5), 1038–1048. doi: 10.1093/brain/awh404

Peñagarikano, O., Lázaro, M. T., Lu, X.-H., Gordon, A., Dong, H., Lam, H. A., et al. (2015). Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. *Sci. Transl. Med.* 7, 271ra8. doi: 10.1126/scitranslmed.3010257

Perry, A., Bentin, S., Shalev, I., Israel, S., Uzefovsky, F., Bar-On, D., et al. (2010). Intranasal oxytocin modulates EEG mu/alpha and beta rhythms during perception of biological motion. *Psychoneuroendocrinology* 35, 1446–1453. doi: 10.1016/j.psyneuen.2010.04.011

Pineda, J. A. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing." *Brain Research. Brain Res. Rev.* 50, 57–68. doi: 10.1016/j.brainresrev.2005.04.005

Prüfer, K., and Jirikowski, G. F. (1997). 1.25-Dihydroxyvitamin D3 receptor is partly colocalized with oxytocin immunoreactivity in neurons of the male rat hypothalamus. *Cell. Mol. Biol.* 43, 543–548.

Puglia, M. H., Lillard, T. S., Morris, J. P., and Connelly, J. J. (2015). Epigenetic modification of the oxytocin receptor gene influences the perception of anger and fear in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3308–3313. doi: 10.1073/pnas.1422096112

Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H.-J., et al. (2014). The role of reward in word learning and its implications for language acquisition. *Curr. Biol.* 24, 2606–2611. doi: 10.1016/j.cub.2014.09.044

Roberts, T. P. L., Khan, S. Y., Rey, M., Monroe, J. F., Cannon, K., Blaskey, L., et al. (2010). MEG detection of delayed auditory evoked responses in autism spectrum disorders: towards an imaging biomarker for autism. *Autism Res.* 3, 8–18. doi: 10.1002/aur.111

Rodenas-Cuadrado, P., Ho, J., and Vernes, S. C. (2014). Shining a light on CNTNAP2: complex functions to complex disorders. *Eur. J. Hum. Genet.* 22, 171–178. doi: 10.1038/ejhg.2013.100

Schaschl, H., Huber, S., Schaefer, K., Windhager, S., Wallner, B., and Fieder, M. (2015). Signatures of positive selection in the cis-regulatory sequences of the human oxytocin receptor (OXTR) and arginine vasopressin receptor 1a (AVPR1A) genes. *BMC Evol. Biol.* 15:85. doi: 10.1186/s12862-015-0372-7

Schonemann, M. D., Ryan, A. K., McEvilly, R. J., O'Connell, S. M., Arias, C. A., Kalla, K. A., et al. (1995). Development and survival of the endocrine hypothalamus and posterior pituitary gland requires the neuronal POU domain factor Brn-2. *Genes Dev.* 9, 3122–3135.

Seltzer, L. J., Ziegler, T. E., and Pollak, S. D. (2010). Social vocalizations can release oxytocin in humans. *Proc. Biol. Sci.* 277, 2661–2666. doi: 10.1098/rspb.2010.0567

Simonyan, K., Horwitz, B., and Jarvis, E. D. (2012). Dopamine regulation of human speech and bird song: a critical review. *Brain Lang.* 122, 142–150. doi: 10.1016/j.bandl.2011.12.009

Singh, F., Nunag, J., Muldoon, G., Cadenhead, K. S., Pineda, J. A., and Feifel, D. (2015). Effects of intranasal oxytocin on neural processing within a socially relevant neural circuit. *Eur. Neuropsychopharmacol.* 26, 626–630. doi: 10.1016/j.euroneuro.2015.12.026

Steinman, G., and Mankuta, D. (2013). Breastfeeding as a possible deterrent to autism–a clinical perspective. *Med. Hypotheses* 81, 999–1001. doi: 10.1016/j.mehy.2013.09.013

Takeda, S., Kuwabara, Y., and Mizuno, M. (1986). Concentrations and origin of oxytocin in breast milk. *Endocrinol. Jpn.* 33, 821–826.

Tamma, R., Colaianni, G., Zhu, L., DiBenedetto, A., Greco, G., Montemurro, G., et al. (2009). Oxytocin is an anabolic bone hormone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7149–7154. doi: 10.1073/pnas.0901890106

Theodosis, D. T. (2002). Oxytocin-secreting neurons: a physiological model of morphological neuronal and glial plasticity in the adult hypothalamus. *Front. Neuroendocrinol.* 23, 101–135. doi: 10.1006/frne.2001.0226

Theofanopoulou, C. (2015). Brain asymmetry in the white matter making and globularity. *Front. Psychol.* 6:1355. doi: 10.3389/fpsyg.2015.01355

Theofanopoulou, C. (2016). "Why oxytocin might affect rhythmicity," in *EVOLANG 11 Workshop: Rhythm: Development, Evolution and Cognition* (New Orleans, LA: Tulane University).

Theofanopoulou, C., and Boeckx, C. (2015). Cognitive phylogenies, the Darwinian logic of descent, and the inadequacy of cladistic thinking. *Front. Cell Dev. Biol.* 3:64. doi: 10.3389/fcell.2015.00064

Uvnäs-Moberg, K., Handlin, L., and Petersson, M. (2015). Self-soothing behaviors with particular reference to oxytocin release induced by non-noxious sensory stimulation. *Front. Psychol.* 5:1529. doi: 10.3389/fpsyg.2014.01529

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., et al. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am. J. Hum. Genet.* 81, 1232–1250. doi: 10.1086/522238

Watanabe, T., Abe, O., Kuwabara, H., Yahata, N., Takano, Y., Iwashiro, N., et al. (2014). Mitigation of sociocommunicational deficits of autism through oxytocin-induced recovery of medial prefrontal activity: a randomized trial. *JAMA Psychiatry* 71, 166–175. doi: 10.1001/jamapsychiatry.2013.3181

Williams, P. G., Dalrymple, N., and Neal, J. (2000). Eating habits of children with autism. *Pediatr. Nurs.* 26, 259–264.

Yamasue, H. (2013). Function and structure in social brain regions can link oxytocin-receptor genes with autistic social behavior. *Brain Dev.* 35, 111–118. doi: 10.1016/j.braindev.2012.08.010

Yau, S. H., McArthur, G., Badcock, N. A., and Brock, J. (2015). Case study: auditory brain responses in a minimally verbal child with autism and cerebral palsy. *Front. Neurosci.* 9:208. doi: 10.3389/fnins.2015.00208

Yoshimura, R., Kimura, T., Watanabe, D., and Kiyama, H. (1996). Differential expression of oxytocin receptor mRNA in the developing rat brain. *Neurosci. Res.* 24, 291–304. doi: 10.1016/0168-010201003-3

Zhao, H., Zhou, W., Yao, Z., Wan, Y., Cao, J., Zhang, L., et al. (2015). Foxp1/2/4 regulate endochondral ossification as a suppresser complex. *Dev. Biol.* 398, 242–254. doi: 10.1016/j.ydbio.2014.12.007

Zheng, J.-J., Li, S.-J., Zhang, X.-D., Miao, W.-Y., Zhang, D., Yao, H., et al. (2014). Oxytocin mediates early experience-dependent cross-modal plasticity in the sensory cortices. *Nat. Neurosci.* 17, 391–399. doi: 10.1038/nn.3634

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

Chapter 3

PROCEEDINGS B

rspb.royalsocietypublishing.org

Review

Cite this article: Theofanopoulou C, Boeckx C, Jarvis ED. 2017 A hypothesis on a role of oxytocin in the social mechanisms of speech and vocal learning. *Proc. R. Soc. B* **284**: 20170988.
http://dx.doi.org/10.1098/rspb.2017.0988

Received: 11 May 2017
Accepted: 20 July 2017

Subject Category:
Neuroscience and cognition

Subject Areas:
neuroscience, evolution

Keywords:
oxytocin, language, vocal learning, birdsong, dopamine

Author for correspondence:
Erich D. Jarvis
e-mail: ejarvis@rockefeller.edu

THE ROYAL SOCIETY
PUBLISHING

A hypothesis on a role of oxytocin in the social mechanisms of speech and vocal learning

Constantina Theofanopoulou[1,3,4], Cedric Boeckx[1,2,3] and Erich D. Jarvis[4,5]

[1]Section of General Linguistics, University of Barcelona, Barcelona, Spain
[2]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain
[3]Universitat de Barcelona Institute for Complex Systems, Barcelona, Spain
[4]Department of Neurogenetics of Language, Rockefeller University, New York, NY, USA
[5]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

CT, 0000-0003-2014-7563; CB, 0000-0001-8882-9718

Language acquisition in humans and song learning in songbirds naturally happen as a social learning experience, providing an excellent opportunity to reveal social motivation and reward mechanisms that boost sensorimotor learning. Our knowledge about the molecules and circuits that control these social mechanisms for vocal learning and language is limited. Here we propose a hypothesis of a role for oxytocin (OT) in the social motivation and evolution of vocal learning and language. Building upon existing evidence, we suggest specific neural pathways and mechanisms through which OT might modulate vocal learning circuits in specific developmental stages.

1. Vocal learning (speech and song) and social experience requirements

Vocal learning is the ability to imitate sounds, found to date in only a few independently evolved species of mammals (humans, bats, cetaceans, sea lions and elephants) and birds (songbirds, parrots and hummingbirds) [1,2]. It is distinct from both auditory and vocal usage learning, which are more ubiquitous among species, and are necessary but not sufficient for vocal learning [3,4].

Early language acquisition is very strongly shaped by social interactions [5]. These social interactions include social motivation for speech learning, emphasized since the dawn of developmental psychology [6]. More recently, social motivation for speech learning has been viewed as a type of social learning [5,7]. Even other forms of sensory–motor learning can involve social feedback [8], and plausibly speech learning could be using a similar mechanism. Several laboratories have experimentally begun to test this hypothesis in humans, and determine to what extent social interactions that modulate attentional, sensory and sensorimotor mechanisms promote language learning. For example, phonological features of babbling are shaped developmentally by social feedback [9] and child speech-related vocalizations (non-cry, non-laugh and non-vegetative) are more likely to receive adults' responses, and in turn, a child's vocalization tends to be speech-related, if the previous speech-related vocalization received an immediate adult feedback [10]. That is, babbling both regulates and is regulated by social interactions, where an infant is socially motivated to learn how to speak, because this learning process is socially rewarding.

This hypothesis, though, needs to be tested with experimental manipulations in non-human animals. The few examples we have from children reared under conditions of social isolation can just partially inform us on the importance of social feedback in language acquisition, both in the auditory and speech domains (as in the famous case of Genie [11]). Kuhl *et al.* [12] managed to tease apart interpersonal interactions from sensory information, by exposing infants to either

audiovisual stimuli or just audio recordings, showing that successful language learning is impossible without social cues. Nonetheless, this experiment was made for second-language learning, leaving unanswered the question of whether the primary speech learning mechanisms can be dissociated from the relevant motivational and rewarding mechanisms provided by social interactions. Further, these experiments tested mainly auditory learning/perception, which is thought to have a different mechanism and brain pathway than speech production learning [4].

Building on the extensive comparative literature on the role of motivation and social decision making in sensory–motor networks [13,14], we think that animal models could shed more light on this issue, and specifically, vocal learning species. Of the non-human vocal learners, songbirds have been studied the most. Songbirds' vocal-learning ability displays parallels to human speech learning, having undergone convergent evolution, at the level of behaviour, neural connectivity and gene expression specializations in song and speech brain regions [2,15,16]. Thus, by dissecting the social mechanisms of vocal learning in songbirds, we could illuminate how social interactions shape vocal learning in humans.

Like human infants, juvenile songbirds learn their songs from social tutors. In laboratory tests, juvenile zebra finches learn best from a live tutor [17]; learning from purely tape recorded songs is less effective, and for some species, often not effective at all [18]. This strong social requirement makes the zebra finch a good candidate for modelling the impact social factors have on human vocal learning. Under conditions between live tutor versus speakers producing song, there are intermediate levels of vocal learning. For example, blindfolded zebra finches interacting with their tutors via grooming or pecking do learn some song, probably in a similar way that blind humans acquire a fully fledged language [17]. Tchernichovski *et al.* [19] have been able to get young zebra finches motivated to learn how to sing without a live tutor, by having them perform an operant conditioning task for the song playback from a fake bird model. When the juveniles have to peck on a key to induce song playbacks from the model, they eagerly keep pecking, and within days to weeks begin to start copying the song from the model [20]. However, if the key is not present and song is played from the model only or the speaker is removed from the model with song played in another location, the juveniles learn very little if at all [21]. These findings indicate that live tutors or fake model birds emit more robust singing social stimuli giving rise to enhanced vocal development, compared to when juvenile songbirds are reared with speakers. This suggests that there could possibly be a social reward mechanism enhancing sensorimotor imitation, a hypothesis that remains to be tested, particularly at the neural and molecular level.

Even though this live versus tape-tutor paradigm could have functioned as the best springboard to study the social mechanisms of vocal learning, researchers have mostly used it to control the auditory parameters the birds get exposed to (e.g. [21]). In addition, since the discovery that male zebra finches alter the structure of their song, gene expression and physiology in song nuclei depending on whether they sing to no one in particular (undirected singing) or to attract a female (directed-singing) [22,23], many studies have focused on adult social interactions after vocal learning is complete. There is a paucity of studies dealing with how social interactions mechanistically affect vocal learning in juvenile songbirds. Among these, Chen *et al.* [24] show that social influences on attention to song enhance vocal learning: tutors altered the structure of their song when directing it to juveniles, reminiscent of the special 'motherese' way humans speak when addressing their speech to infants.

Deciphering the mechanisms of the social motivation of vocal learning, and determining whether the mechanism of social motivation to learn vocalizations can be dissociated from the act of vocal learning, we believe requires figuring out the circuit and molecular mechanisms. Towards this end, we propose that the neuropeptide oxytocin (OT) and its social reward circuitry make a very good candidate that could control the social reward mechanisms for vocal learning.

2. Oxytocin as a good candidate to control social motivation of vocal learning

Oxytocin, depending on the brain region and release site, acts as a hormone, neuromodulator or neurotransmitter that functions through its receptor (OTR) to regulate a diverse set of biological processes: pregnancy and uterine contractions, milk ejection, attachment between mothers and their young, bond formation, copulation and orgasm, suppression of stress, thermoregulation, olfactory processing, eye contact and recognition of familiar individuals [25], with the caveat that some functions are specific to one lineage, such as mammals. OT is thought to have its effect on many systems because it is most prominently expressed in hypothalamic OT neurons that project to many brain regions where the receptor is located [26,27]. Recent studies attest that OT enhances socially reinforced learning in humans and rhesus macaques [28,29], while other studies show its involvement in vocal and auditory behaviours (see references herein). As a result, Theofanopoulou [30] put forth the hypothesis that OT might be implicated in cognitive aspects of language processing in humans. Here we adduce more evidence also for a role in the social motivation of language learning. We further sketch out possible mechanisms for social motivation for vocal learning in vocal-learning species. With regard to gene terminology, we have adopted a universal nomenclature based on sequence identity and gene synteny, using the same gene name OT and OTR across vertebrates [31].

(a) Vocal non-learners

We first note that OT appears to have a role in auditory–vocal communication even before vocal learning evolved, as such a role can be found in vocal non-learning species that span the vertebrate phylogeny, from fish to mammals. Goodson & Bass [32] found that OT in midshipman fish modulates the burst duration in the innate vocalizations that sneaker males and females produce in non-reproductive contexts. OT immunoreactive cell groups are distributed throughout their vocal–acoustic circuit, from the midbrain to the forebrain [33]. In rats, OT enhances both inhibitory and excitatory synaptic currents in the hypoglossal motor nucleus which innervates the tongue muscles, thus potentially controlling rat vocalizations [34]. In mice, Winslow *et al.* [35] found that infant OT-KO (knock-out) animals were less vocal than wild-type (WT) controls during separations from the mother and peers. Likewise, Takayanagi *et al.* [36] observed fewer ultrasonic vocalizations emitted by infant OTR-KO compared with wild-type mice in a social isolation paradigm. Marlin *et al.* [37] demonstrated that when inexperienced virgin females

are given OT intraperitoneally or through optogenetic stimulation of hypothalamic OT neuronal axons that project into the auditory cortex, and then co-housed with a mother and her litter, their retrieval of vocalizing pups was effective as if they were the mother. Follow-up studies showed that OTR levels are remarkably lateralized with higher expression in neurons of the left auditory cortex [26]. These observations in vocal non-learners strike us as particularly relevant, as auditory and vocal learning/language circuits in humans are mainly left-lateralized [38], and are either left- or right-lateralized among different species of song-learning birds [39].

Based on these findings, we suggest that OT has a role in social motivation of auditory and vocal communication behaviours in vocal non-learners, and that a lateralized function in the auditory cortex may have been present before vocal learning and language evolved. Even though less likely, it could still be possible that OT influences the social motivation for vocal learning through the OTR receptors in innate brainstem circuits, including two auditory brainstem nuclei, the nucleus magnocellularis (NM) and the nucleus laminaris (NL) and in the vocal motor neurons (nXIIts) in songbirds [40].

(b) Vocal learners

In humans, intranasal OT administration modulates semantic integration in speech comprehension [41]. In autistic patients, the oxytocinergic system has been repeatedly indicated to function aberrantly. Specifically, Rijlaarsdam *et al.* [42] identified a significant OXTR rs53576 genotype by OXTR methylation interaction associated with communication problems in autistic patients, while Zhang *et al.* [43] found that autistic children with higher plasma OT concentrations tended to have less impairment of verbal communication. In turn, after OT intranasal administration, autistic patients had a more efficient and long-lasting performance in a speech comprehension task [44,45]. Based on findings that intranasal administration of OT crosses the blood–brain barrier and binds to areas where the receptors are located [46], we can interpret these studies as bearing directly on our hypothesis.

In songbirds, experimental manipulation of the oxytocinergic system with OT agonist and antagonist have been made mostly in the context of pair-bonding and aggression, with very few and some controversial reports on how these treatments affected singing, probably due to different treatment sites [47–49]. Nevertheless, OT has been found to affect the amount of directed singing to females [48]. These findings in vocal learning species indicate that OT may also have a social enhancement for aspects of auditory processing and learned vocal communication.

(c) Neural pathways

In order for our hypothesis to have some validity, OT would be expected to innervate vocal learning circuits directly, that in turn would express the OTR, or indirectly via other motivation/reward circuits that, in turn, innervate vocal learning circuits [50]. All vocal learning species examined to date (humans and the song learning birds) have a highly specialized forebrain circuit that controls learning and production of learned sounds (figure 1*a*) [4,16]. Best studied in songbirds, the pathway consists of an anterior forebrain circuit that controls vocal imitation and a posterior circuit that controls production of learned vocalizations. The anterior forebrain circuit consists of LMAN in the cortical region, Area X in the striatum and aDLM in the thalamus, which form a pallial-basal ganglia-thalamo-pallial loop (figure 1*a*). When Area X is lesioned in juveniles, the birds are not able to crystallize onto a learned song, as their vocalizations remain variable. Conversely, when LMAN is lesioned, the bird instantly crystallizes onto what it had learned up to that moment [51]. These and other findings lead to one interpretation being that during the juvenile vocal learning period, Area X injects stereotypy, whereas LMAN injects variability into the vocalizations, and the two opposing functions enable vocal imitation [2,51]. After learning is complete, lesions in adults, such as in Area X, lead to deficits in song sequencing (or production) similar to stuttering in humans [52,53]. The posterior pathway in songbirds consists of the HVC and RA, thought to control sequencing and acoustic structure of syllables, respectively. In humans, the analogous anterior pathway has been proposed to be a cortical-basal ganglia-thalamo-pallial loop involving Broca's area (LMAN analogue), part of the anterior striatum (ASt) and the anterior thalamus; the analogous human posterior pathway has been proposed to include the laryngeal motor cortex (LMC; figure 1*a*), with different cortical layers representing songbird HVC (layer 3) and RA (layer 5) [2,16]. This forebrain vocal pathway is either absent or limited at best in vocal non-learning species, including non-human primates and mice [54,55]. But all vocal learning and non-learning species have a more comparable auditory forebrain pathway, involved in auditory learning, as described above for the mouse pup retrieval experiments.

The OTR is expressed broadly across cortical and subcortical brain regions, in both mammals and birds, including humans [27,56,57]. However, in different species there are brain regions with enriched OTR expression relative to all other brain regions, and they often correlate with differences in social behaviours between species [58]. We are not aware of anyone determining if there is enriched specialized expression (increased or decreased relative to adjacent brain regions) in speech brain regions in humans. In songbirds, some limited expression analyses in the posterior pathway revealed differences between species, with specialized upregulation of OTR in HVC and downregulation in the RA compared with the surrounding motor regions in zebra finches, that sing one simple song and higher expression (although not shown) in white throated sparrows, a species that sings at least two different songs [40]. One prediction from these findings would be that human LMC layer 5 neurons may have downregulation of OTR relative to layer 5 neurons of the adjacent non-speech cortex. We noted from our examination of fig. 1 of [58] that there is layer-specific expression of OTR in the motor cortex that is different across rodent species. We also predict some within-species differences, such as in songbirds where females lost the vocal learning trait [59] and would not be expected to have forebrain vocal OT neuron innervation.

In terms of possible indirect interactions through other reward brain circuits, hypothalamic OT neurons innervate the ventral tegmental area (VTA), which innervates the vocal learning systems in humans and songbirds [60,61]. The VTA releases dopamine (DA) mainly to striatal brain regions and some cortical brain regions, including vocal learning regions in songbirds [62]; through its DA receptors it is thought to reinforce learning and motivated behaviour. There is a plethora of evidence in the mammalian literature showing that OT neurons in the hypothalamic paraventricular and supraoptic nuclei (PVN and SON) send projections to the VTA, and

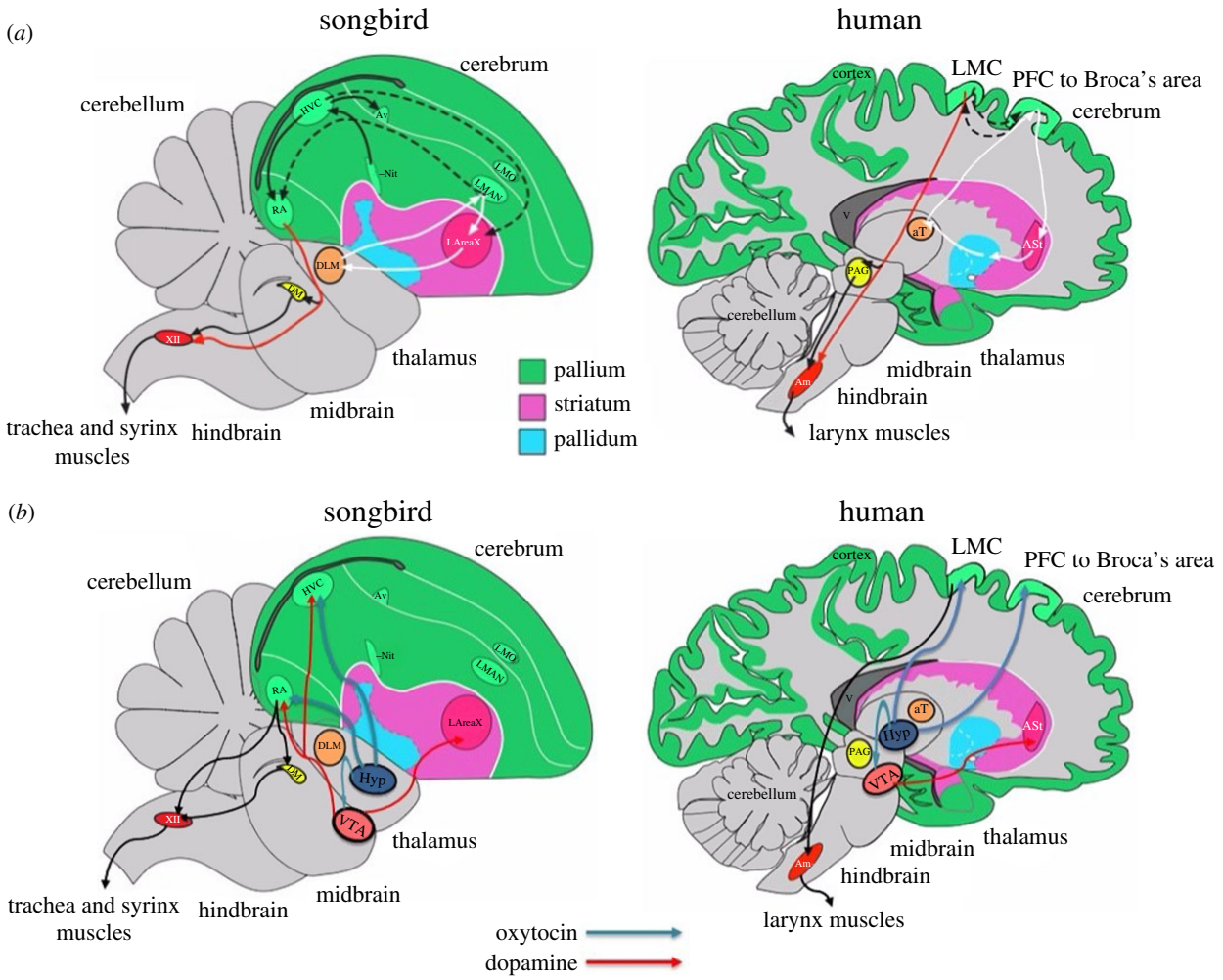


Figure 1. Summary diagrams of vocal learning systems in songbirds and humans. (*a*) Vocal learning circuits. Red arrows, the direct posterior forebrain projection to vocal motor neurons in the brainstem. White lines, anterior forebrain circuit. Dashed lines, connections between the anterior and posterior vocal motor circuits. (*b*) Proposed oxytocinergic and dopaminergic projections into the vocal learning circuits. In songbirds, we propose oxytocinergic neurons from the Hyp project to the RA, HVC and VTA; VTA makes a strong dopaminergic projection to LAreaX and weaker ones to HVC and RA. In humans, we propose oxytocinergic neurons from the Hyp project to the LMC, Broca's area and the VTA; VTA makes dopaminergic projections to the ASt. Black arrows, connectivity of the proposed system with the brainstem. Abbreviations: HVC, HVC nucleus; LMAN, lateral magnocellular nucleus of anterior nidopallium; RA, robust nucleus of arcopallium; Area X, area X of the striatum; Hyp, hypothalamus; VTA, ventral tegmental area; DLM, dorsal lateral nucleus of the medial thalamus; Av, nucleus avalanche; LMO, lateral oval nucleus of the mesopallium; NIf, interfacial nucleus of the nidopallium; DM, dorsal medial nucleus of the midbrain; XII, 12th nucleus, tracheosyringeal part; PFC, prefrontal cortex; LMC, laryngeal motor cortex; A St, anterior striatum; PAG -periaqueductal grey; aT, anterior thalamus; Am, nucleus ambiguus of the brainstem. Note: The position of Broca's area is shown here more medially for simplicity. (Adapted from [4,16].)

stimulate DA neurons there [63–66]. Consistent with this, in the last several decades, a number of studies have shown OT–DA interactions in many social behaviours [67,68]. The VTA expresses OTR [69,70] and injection of OT into the VTA of rats increases DA release, inducing penile erection [63,71,72]. Intracerebroventricular injection of an OTR antagonist attenuates DA agonist-stimulated DA release and the pro-erectile effect [64]. Peris *et al.* [73] infected mice with a Cre-inducible adeno-associated virus that drives the expression of an OTR-fluorescent reporter in the VTA and found that OTR-expressing neurons in VTA project to the nucleus accumbens, prefrontal cortex, the extended amygdala and other forebrain regions; also some of these neurons were identified as DA neurons. Bromberg-Martin *et al.* [74] have also shown that DA neurons within the VTA encode motivationally salient signals. Thus, OT, by modulating activity within the DA system, may alter the assignment of motivational salience.

Lastly, since OT can bind to one of the vasopressin/vasotocin receptors (vasopressin/vasotocin receptor 1A; AVPR1A or V1AR) with equal affinity as it does to the OTR [75], we do not exclude the possibility that OT may be playing a role in the social motivation for vocal learning via this receptor too. It is also the only vasopressin/vasotocin receptor thus far found to be expressed in vocal learning regions [40] and to be involved in singing behaviour [76].

Taking the behavioural, circuit and molecular findings together, we suggest that OT and the circuits it functions in are good candidates for the long-hypothesized motivation and reward mechanism of vocal learning. Part of the mechanisms may have been present before vocal learning evolved, but part of it may be specialized in vocal learning circuits and behaviour. With this information, we propose a testable mechanism, either via direct influence on vocal learning pathways or indirect through the VTA DA neuron pathway.

3. Proposed neural and molecular mechanisms

In this section, we consider the *when*, *where* and *how* OT might modulate socially motivated vocal learning behaviour.

For the *when*, we consider the three major stages vocal learners are known to acquire their ability to imitate vocalizations: sensory, sensorimotor practice and crystallization. In some species these stages can be distinctly separate, and in others they overlap. In the first sensory infant/nestling phase, vocal learning animals and humans acquire auditory memories of the vocalizations that they hear through social interactions [15]. In this phase, it is not necessary that the animal or child imitate or even vocalize. In the second, sensorimotor phase, as in other cases of sensorimotor learning, vocal learning proceeds through a reinforcement learning mechanism [50,61], where juvenile song-learning birds begin to produce semi-imitated vocalizations, and evaluate their own motor output via sensory feedback and reinforce it only if it closely matched the predicted outcome [77]. A mechanism proposed for reinforcement is the variability observed in juvenile song, suggestive of a motor exploration [78], with reinforcement (or error) neural signals guiding song imitation [79]. Likewise, human infants in this phase appear to experiment with uttering articulate sounds, but without yet producing recognizable words (i.e. babbling). Again reward behaviours, e.g. when parents complement their child with excited words, clapping, smiles and hugs, after their first speech-related attempts, reinforces speech learning. In the third crystallization phase, as they become adults (i.e. puberty phase in humans), song-learning birds and speech-learning humans complete the development of their vocal repertoire, and the ability to learn new vocalizations/languages is either shut down (e.g. zebra finches) or made more difficult (e.g. canaries and humans). However, if a song-learning bird is removed from its conspecifics before this phase is complete, it will take significantly longer for the animal to crystallize on a repertoire [80]. We propose that OT will have its effects during the sensory and sensorimotor phases of vocal learning, and less so during or post crystallization, because the first two phases are more dependent on social experience. It is also likely that the same mechanisms could apply throughout life, but at a more reduced level.

For the *where* and *how*, during the sensory phase, we propose that OT could enhance the formation of socially driven auditory memories that impinge on the vocal learning circuit. This could occur by a direct projection of OT hypothalamic neurons into the auditory cortex, as seen in adult female mice for pup retrieval, or by direct projections to the vocal learning pathway brain regions. For the former possibility, auditory input into juvenile HVC from a playback has been found to modulate its neural connectivity and function in song production [81]. We propose that when a vocal learning infant/juvenile hears vocalizations generated from a conspecific, there could be an associated increase in OT release into the auditory and/or specialized vocal learning brain regions to strengthen the newly formed synaptic interactions to hold onto the memories and shape the vocal learning pathway. Similar to DA circuits (see below), this strengthening and shaping could occur by OT binding to OTR in neurons of the auditory and vocal pathways that receive excitatory and inhibitory inputs for the auditory–vocal memories. A prediction of this hypothesis is that if the auditory signals are not from a social individual, the auditory processing circuits would still process the sounds and form auditory memories of them, but the OT circuit would not strengthen the auditory input to the vocal learning circuit for eventual imitation of the sounds.

During the sensorimotor learning phase, we propose again that OT input to the auditory and vocal learning pathways could be activated, but this time by positive social feedback (auditorily or by other means) from conspecifics when the juvenile produces more accurate copies of the learned vocalizations. The positive feedback could help strengthen the connections that control production of the more accurate copy of the vocalizations. But could OT also modulate imitation of vocalizations during sensorimotor practice independent of immediate social input from others? Although this would move us away from a direct social role of OT in vocal learning, we consider the possibility that self-motivation and even purely vocal learning mechanisms independent of immediate social mechanisms could also be involved. For this possibility, we turn to studies on DA.

As described earlier, there is a robust VTA DA-neuron projection to vocal learning nucleus Area X (figure 1*b*) [60,62]. An analogous vocal learning region has been found in the human striatum, with many of same gene expression specializations as in songbird Area X [16,82]. VTA also makes a weaker, but still relatively prominent, projection to the vocal production nuclei HVC and RA, and receives input from an auditory area around RA necessary for vocal learning [83]. DA levels in Area X are higher during directed singing (to females) than undirected singing, due to differential activity of the re-uptake transporter (a noradrenaline transporter in birds), in the VTA axons within Area X [84]. When this transporter is pharmacologically blocked, DA levels during undirected singing reach the levels of DA release during directed singing [84]. Unilateral lesions of the VTA dopaminergic projections reduce singing-driven Immediate Early Gene (IEG) expression in Area X in both contexts [60]. More recently, Gadagkar *et al.* [79] showed that the VTA DA neurons that project to Area X encode performance error-and-reward during singing, where these neurons are suppressed when the bird simultaneously hears distorted feedback syllables and are activated when they hear undistorted syllables. It is plausible to hypothesize that such *performance* signals might subserve vocal *learning* in juvenile animals, when the songbird monitors if the vocal output produced matched 'the desired tutor outcome, and also the predicted probability of achieving the desired outcome' [79]. Recently, Chen *et al.* [24] found that in juvenile animals the percentage of DA neurons expressing EGR-1 (an IEG) in the VTA was significantly higher in socially tutored juveniles relative to passively tutored juveniles with playbacks of songs from a speaker or untutored juveniles, indicating that this neural correlate might be responsible for the differences in vocal-learning performance.

We propose that OT might have a role in both the social motivation and the sensorimotor mechanisms of vocal learning via hypothalamic OT action on VTA DA-neurons that project to Area X and other song nuclei (figure 1*b*). During sensorimotor practice, there could be self-induced motivation of the OT → VTA → song nuclei circuits to help strengthen connections within the circuit when the tutee's produced song matches his auditory memory of the tutor's song. After vocal learning is complete, the presumed downregulation of OTR in several vocal learning nuclei (relative to the surrounding brain regions) in zebra finches may contribute to crystallization and shutting off the ability to further imitate from conspecifics. For the latter part of the hypothesis to be plausible, one would need to determine if there are higher levels of OTR in these song nuclei during juvenile development.

It is important to mention that up to now the only well-studied hypothesis of where the VTA gets its input for vocal

learning functions has been articulated by Riters [85] and colleagues based on their studies on European starlings. According to them, it is the projection from the medial preoptic nucleus (mPOA) to the VTA that is crucial for social reward-related functions of vocal learning. Our hypothesis shifts the focus from the mPOA to the hypothalamus, because of the OTR expression in song-learning nuclei and suggestive direct influence of hypothalamic OT on singing. The two hypotheses could be complementary: OTRs are also expressed in the mPOA and mPOA neurons (at least in mammals) that project to the VTA (or to the PVN and from there to VTA) [86] play a role in social bonding regulation and maintenance [72]. That is, the OT input to the VTA could be originating both from the hypothalamus (PVN/SON → VTA) and the mPOA (mPOA → VTA or mPOA → PVN → VTA).

In humans, building on OTR expression patterns in the brain [27,56,57], we propose that OT neurons might project directly from the hypothalamus to the LMC and Broca's area or indirectly to them and other speech-regions through the VTA (figure 1*b*). Regarding the latter, there is evidence that OT administration enhances activation in the VTA of humans [87]. In this manner, OT might affect VTA's DA output to the anterior striatum speech region [88] and LMC [54,61], and from there (LMC) to the vocal motor nucleus ambiguous of the brainstem. Given these similar findings in humans, we see no reason to propose a fundamentally different mechanism for the sensory and sensorimotor learning phases of vocal learning in humans or other vocal learning species.

An alternative route through which OT could also affect the social motivation for vocal learning is through its hormonal action via the hypothalamic–pituitary–adrenal axis, known for attenuating the stress response [89] and thus making social learning more efficient. However, we deem this possibility as less likely, given that the OTR is found in the auditory cortex and in speech/song areas, most likely directly affecting vocal learning.

4. Proposed experiments to test hypothesis

In this final section, we offer some proposed experiments that would validate or falsify some of the key tenants of our hypothesis.

A prediction of our hypothesis that OT controls the social motivation to imitate vocalizations in vocal learners, is that blocking OT in the brain, and more specifically its targets to the OTR in auditory cortex, vocal learning neurons and/or VTA-DA neurons during the sensory and sensorimotor phases would prevent vocal learning from live social or model tutors. Conversely, activating OT in these circuits, when a young juvenile hears novel vocalizations from a live tutor or a tape recorder, would potentially cause the juvenile to imitate the song heard better and also treat that tape recorder as more of a social object. This would also mean that OT neuron activation and release, and activation of OTR in the target brain regions, would also change in the same direction. It is not feasible or ethical to conduct such experiments in humans, but they can be conducted in a non-human vocal learning species, such as songbirds.

Because our hypothesis is at its infant stage, informing and testing the hypothesis further will also require a great amount of more descriptive research. This includes: (i) a detailed expression analyses of OT, OTR and associated family of genes (vasopressin/vasotocin and its receptors) in the vocal communication brain regions throughout development and adulthood, across multiple vocal learning and non-learning species; (ii) analyses of coding sequence and regulatory regions of these genes to determine if there are convergent genetic changes in vocal learning species that could explain brain functional or expression differences, respectively and (iii) physiology analyses of OT neurons and OT release during vocal learning and language acquisition. Some of these more descriptive experiments can be done with humans and non-human primates, and thus offer a more direct window to inform our hypothesis on OT function in language.

5. Conclusion

We have sketched out what we consider a plausible hypothesis of a role for OT in the social motivation of vocal learning and language. This hypothesis, if validated, would fill in a gap in our knowledge of the main molecule(s) that control the social motivation for vocal learning. With this hypothesis, we are able to assemble disparate pieces of knowledge into a greater whole, with OT as a nexus. As in all hypotheses, there are parts that have weaknesses in ours, such as whether OT modulation of vocal learning circuits and thus language are direct or indirect. For these, we propose plausible alternative mechanisms that can be tested and modified with new knowledge. Overall, though, we find it hard to come up with a better viable alternative hypothesis, given the current state of knowledge. Thus, we believe the hypothesis we propose at this time is the most attractive one worth testing.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. C.T. acknowledges support from the Generalitat de Catalunya. C.B. acknowledges the financial support from the Spanish Ministry of Economy and Competitiveness (grant FFI2016-78034-C2-1-P), a Marie Curie International Reintegration grant (PIRG-GA-2009-256413), research funds from the Fundació Bosch i Gimpera and from the Generalitat de Catalunya (2014-SGR-200). E.D.J. was funded by the Howard Hughes Medical Institute and the Rockefeller University.

References

1. Janik VM, Slater PJ. 1997 Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–100. (doi:10.1016/S0065-3454(08)60377-0)
2. Jarvis ED. 2004 Learned birdsong and the neurobiology of human language. *Ann. NY Acad. Sci.* **1016**, 749–777. (doi:10.1196/annals.1298.038)
3. Janik VM, Slater PJ. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11. (doi:10.1006/anbe.2000.1410)
4. Petkov CI, Jarvis E. 2012 Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12. (doi:10.3389/fnevo.2012.00012)
5. Kuhl P. 2007 Is speech learning 'gated' by the social brain? *Dev. Sci.* **10**, 110–120. (doi:10.1111/j.1467-7687.2007.00572.x)
6. Bruner J. 1985 The role of interaction formats in language acquisition. In *Language and social situations*, pp. 31–46. Berlin, Germany: Springer.

7. Tomasello M. 2009 *Constructing a language*. Harvard, UK: Harvard University Press.
8. Galea JM, Mallia E, Rothwell J, Diedrichsen J. 2015 The dissociable effects of punishment and reward on motor learning. *Nat. Neurosci.* **18**, 597–602. (doi:10.1038/nn.3956)
9. Goldstein MH, King AP, West MJ. 2003 Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl Acad. Sci. USA* **100**, 8030–8035. (doi:10.1073/pnas.1332441100)
10. Warlaumont AS, Richards JA, Gilkerson J, Oller DK. 2014 A social feedback loop for speech development and its reduction in autism. *Psychol. Sci.* **25**, 1314–1324. (doi:10.1177/0956797614531023)
11. Fromkin V, Krashen S, Curtiss S, Rigler D, Rigler M. 1974 The development of language in genie: a case of language acquisition beyond the 'critical period'. *Brain Lang.* **1**, 81–107. (doi:10.1016/0093-934X(74)90027-3)
12. Kuhl PK, Tsao F-M, Liu H-M. 2003 Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl Acad. Sci. USA* **100**, 9096–9101. (doi:10.1073/pnas.1532872100)
13. O'connell LA, Hofmann HA. 2011 The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. *J. Comp. Neurol.* **519**, 3599–3639. (doi:10.1002/cne.22735)
14. Syal S, Finlay BL. 2011 Thinking outside the cortex: social motivation in the evolution and development of language. *Dev. Sci.* **14**, 417–430. (doi:10.1111/j.1467-7687.2010.00997.x)
15. Doupe AJ, Kuhl PK. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
16. Pfenning AR *et al*. 2014 Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846. (doi:10.1126/science.1256846)
17. Eales LA. 1989 The influences of visual and vocal interaction on song learning in zebra finches. *Anim. Behav.* **37**, 507–508. (doi:10.1016/0003-3472(89)90097-3)
18. Beecher MD. 2016 Birdsong learning as a social process. *Anim. Behav.* **124**, 233–246. (doi:10.1016/j.anbehav.2016.09.001)
19. Tchernichovski O, Mitra PP, Lints T, Nottebohm F. 2001 Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291**, 2564–2569. (doi:10.1126/science.1058522)
20. Deshpande M, Pirlepesov F, Lints T. 2014 Rapid encoding of an internal model for imitative learning. *Proc. R. Soc. B* **281**, 20132630. (doi:10.1098/rspb.2013.2630)
21. Derégnaucourt S, Poirier C, Van der Kant A, Van der Linden A, Gahr M. 2013 Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *J. Physiol. Paris* **107**, 210–218. (doi:10.1016/j.jphysparis.2012.08.003)
22. Jarvis ED, Scharff C, Grossman MR, Ramos JA, Nottebohm F. 1998 For whom the bird sings: context-dependent gene expression. *Neuron* **21**, 775–788. (doi:10.1016/S0896-6273(00)80594-2)
23. Hessler NA, Doupe AJ. 1999 Social context modulates singing-related neural activity in the songbird forebrain. *Nat. Neurosci.* **2**, 209–211. (doi:10.1038/6306)
24. Chen Y, Matheson LE, Sakata JT. 2016 Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc. Natl Acad. Sci. USA* **113**, 6641–6646. (doi:10.1073/pnas.1522306113)
25. Choleris E, Pfaff DW, Kavaliers M. 2013 *Oxytocin, vasopressin and related peptides in the regulation of behavior*. Cambridge, UK: Cambridge University Press.
26. Mitre M, Marlin BJ, Schiavo JK, Morina E, Norden SE, Hackett TA, Aoki CJ, Chao MV, Froemke RC. 2016 A distributed network for social cognition enriched for oxytocin receptors. *J. Neurosci.* **36**, 2517–2535. (doi:10.1523/JNEUROSCI.2409-15.2016)
27. Boccia M, Petrusz P, Suzuki K, Marson L, Pedersen CA. 2013 Immunohistochemical localization of oxytocin receptors in human brain. *Neuroscience* **253**, 155–164. (doi:10.1016/j.neuroscience.2013.08.048)
28. Clark-Elford R. 2014 The effects of oxytocin on social reward learning in humans. *Int. J. Neuropsychopharmacol.* **17**, 199–209. (doi:10.1017/S1461145713001120)
29. Parr LA. 2014 Intranasal oxytocin enhances socially-reinforced learning in rhesus monkeys. *Front. Behav. Neurosci.* **8**, 278. (doi:10.3389/fnbeh.2014.00278)
30. Theofanopoulou C. 2016 Implications of oxytocin in human linguistic cognition: from genome to phenome. *Front. Neurosci.* **10**, 271. (doi:10.3389/fnins.2016.00271)
31. Lagman D, Daza DO, Widmark J, Abalo XM, Sundström G, Larhammar D. 2013 The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. *BMC. Evol. Biol.* **13**, 238. (doi:10.1186/1471-2148-13-238)
32. Goodson JL, Bass AH. 2000 Forebrain peptides modulate sexually polymorphic vocal circuitry. *Nature* **403**, 769–772. (doi:10.1038/35001581)
33. Goodson JL, Evans AK, Bass AH. 2003 Putative isotocin distributions in sonic fish: relation to vasotocin and vocal–acoustic circuitry. *J. Comp. Neurol.* **462**, 1–14. (doi:10.1002/cne.10679)
34. Wrobel L, Reymond-Marron I, Dupré A, Raggenbass M. 2010 Oxytocin and vasopressin enhance synaptic transmission in the hypoglossal motor nucleus of young rats by acting on distinct receptor types. *Neuroscience* **165**, 723–735. (doi:10.1016/j.neuroscience.2009.11.001)
35. Winslow JT, Hearn EF, Ferguson J, Young LJ, Matzuk MM, Insel TR. 2000 Infant vocalization, adult aggression, and fear behavior of an oxytocin null mutant mouse. *Horm. Behav.* **37**, 145–155. (doi:10.1006/hbeh.1999.1566)
36. Takayanagi Y *et al*. 2005 Pervasive social deficits, but normal parturition, in oxytocin receptor-deficient mice. *Proc. Natl Acad. Sci. USA* **102**, 16 096–16 101. (doi:10.1073/pnas.0505312102)
37. Marlin BJ, Mitre M, D'amour JA, Chao MV, Froemke RC. 2015 Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* **520**, 499–504. (doi:10.1038/nature14402)
38. Galaburda AM, Sanides F, Geschwind N. 1978 Human brain: cytoarchitectonic left–right asymmetries in the temporal speech region. *Arch. Neurol.* **35**, 812–817. (doi:10.1001/archneur.1978.00500360036007)
39. Voss HU, Tabelow K, Polzehl J, Tchernichovski O, Maul KK, Salgado-Commissariat D, Ballon D, Helekar SA. 2007 Functional MRI of the zebra finch brain during song stimulation suggests a lateralized response topography. *Proc. Natl Acad. Sci. USA* **104**, 10 667–10 672. (doi:10.1073/pnas.0611515104)
40. Leung CH, Abebe DF, Earp SE, Goode CT, Grozhik AV, Mididoddi P, Maney DL. 2011 Neural distribution of vasotocin receptor MRNA in two species of songbird. *Endocrinology* **152**, 4865–4881. (doi:10.1210/en.2011-1394)
41. Ye Z, Stolk A, Toni I, Hagoort P. 2016 Oxytocin modulates semantic integration in speech comprehension. *J. Cogn. Neurosci.* **29**, 267–276. (doi:10.1162/jocn_a_01044)
42. Rijlaarsdam J, van IJzendoorn MH, Verhulst FC, Jaddoe VW, Felix JF, Tiemeier H, Bakermans-Kranenburg MJ. 2016 Prenatal stress exposure, oxytocin receptor gene (*OXTR*) methylation and child autistic traits: the moderating role of *OXTR* rs53576 genotype. *Autism Res.* **10**, 430–438. (doi:10.1002/aur.1681)
43. Zhang H-F, Dai Y-C, Wu J, Jia M-X, Zhang J-S, Shou X-J, Han S-P, Zhang R, Han J-S. 2016 Plasma oxytocin and arginine-vasopressin levels in children with autism spectrum disorder in China: associations with symptoms. *Neurosci. Bull.* **32**, 423–432. (doi:10.1007/s12264-016-0046-5)
44. Hollander E, Bartz J, Chaplin W, Phillips A, Sumner J, Soorya L, Anagnostou E, Wasserman S. 2007 Oxytocin increases retention of social cognition in autism. *Biol. Psychiatry* **61**, 498–503. (doi:10.1016/j.biopsych.2006.05.030)
45. Pfundmair M, Lamprecht F, von Wedemeyer FM, Frey D. 2016 Your word is my command: oxytocin facilitates the understanding of appeal in verbal communication. *Psychoneuroendocrinology* **73**, 63–66. (doi:10.1016/j.psyneuen.2016.07.213)
46. Lee M, Scheidweiler K, Diao X, Akhlaghi F, Cummins A, Huestis M, Leggio L, Averbeck B. 2017 Oxytocin by intranasal and intravenous routes reaches the cerebrospinal fluid in rhesus macaques: determination using a novel oxytocin assay. *Mol. Psychiatry*. (doi:10.1038/mp.2017.27)
47. Goodson JL, Lindberg L, Johnson P. 2004 Effects of central vasotocin and mesotocin manipulations on social behavior in male and female zebra finches. *Horm. Behav.* **45**, 136–143. (doi:10.1016/j.yhbeh.2003.08.006)
48. Pedersen A, Tomaszycki M. 2012 Oxytocin antagonist treatments alter the formation of pair relationships in zebra finches of both sexes. *Horm. Behav.* **62**, 113–119. (doi:10.1016/j.yhbeh.2012.05.009)
49. Klatt JD, Goodson JL. 2013 Oxytocin-like receptors mediate pair bonding in a socially monogamous

songbird. *Proc. R. Soc. B* **280**, 20122396. (doi:10.1098/rspb.2012.2396)
50. Fee MS, Goldberg JH. 2011 A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* **198**, 152–170. (doi:10.1016/j.neuroscience.2011.09.069)
51. Scharff C, Nottebohm F. 1991 A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11**, 2896–2913.
52. Kobayashi K, Uno H, Okanoya K. 2001 Partial lesions in the anterior forebrain pathway affect song production in adult Bengalese finches. *Neuroreport* **12**, 353–358. (doi:10.1097/00001756-200102120-00034)
53. Kubikova L, Turner EA, Jarvis ED. 2007 The pallial basal ganglia pathway modulates the behaviorally driven gene expression of the motor pathway. *Eur. J. Neurosci.* **25**, 2145–2160. (doi:10.1111/j.1460-9568.2007.05368.x)
54. Jurgens U. 1982 Afferents to the cortical larynx area in the monkey. *Brain Res.* **239**, 377–389. (doi:10.1016/0006-8993(82)90516-9)
55. Arriaga G, Jarvis ED. 2013 Mouse vocal communication system: are ultrasounds learned or innate? *Brain Lang.* **124**, 96–116. (doi:10.1016/j.bandl.2012.10.002)
56. Loup F, Tribollet E, Dubois-Dauphin M, Dreifuss J. 1991 Localization of high-affinity binding sites for oxytocin and vasopressin in the human brain. an autoradiographic study. *Brain Res.* **555**, 220–232. (doi:10.1016/0006-8993(91)90345-V)
57. Bethlehem R *et al*. 2017 Intranasal oxytocin enhances intrinsic corticostriatal functional connectivity in women. *Transl. Psychiatry* **7**, e1099. (doi:10.1038/tp.2017.72)
58. Anacker AM, Beery AK. 2013 Life in groups: the roles of oxytocin in mammalian sociality. *Front Behav. Neurosci.* **7**, 185. (doi:10.3389/fnbeh.2013.00185)
59. Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE. 2014 Female song is widespread and ancestral in songbirds. *Nat. Commun.* **5**, 3379. (doi:10.1038/ncomms4379)
60. Hara E, Kubikova L, Hessler NA, Jarvis ED. 2007 Role of the midbrain dopaminergic system in modulation of vocal brain activation by social context. *Eur. J. Neurosci.* **25**, 3406–3416. (doi:10.1111/j.1460-9568.2007.05600.x)
61. Simonyan K, Horwitz B, Jarvis ED. 2012 Dopamine regulation of human speech and bird song: a critical review. *Brain Lang.* **122**, 142–150. (doi:10.1016/j.bandl.2011.12.009)
62. Lewis JW, Ryan SM, Arnold AP, Butcher LL. 1981 Evidence for a catecholaminergic projection to area X in the zebra finch. *J. Comp. Neurol.* **196**, 347–354. (doi:10.1002/cne.901960212)
63. Melis MR, Melis T, Cocco C, Succu S, Sanna F, Pillolla G, Boi A, Ferri G-L, Argiolas A. 2007 Oxytocin injected into the ventral tegmental area induces penile erection and increases extracellular dopamine in the nucleus accumbens and paraventricular nucleus of the hypothalamus of male rats. *Eur. J. Neurosci.* **26**, 1026–1035. (doi:10.1111/j.1460-9568.2007.05721.x)
64. Succu S, Sanna F, Melis T, Boi A, Argiolas A, Melis MR. 2007 Stimulation of dopamine receptors in the paraventricular nucleus of the hypothalamus of male rats induces penile erection and increases extra-cellular dopamine in the nucleus accumbens: involvement of central oxytocin. *Neuropharmacology* **52**, 1034–1043. (doi:10.1016/j.neuropharm.2006.10.019)
65. Melis MR, Succu S, Sanna F, Boi A, Argiolas A. 2009 Oxytocin injected into the ventral subiculum or the posteromedial cortical nucleus of the amygdala induces penile erection and increases extracellular dopamine levels in the nucleus accumbens of male rats. *Eur. J. Neurosci.* **30**, 1349–1357. (doi:10.1111/j.1460-9568.2009.06912.x)
66. Succu S, Sanna F, Argiolas A, Melis MR. 2011 Oxytocin injected into the hippocampal ventral subiculum induces penile erection in male rats by increasing glutamatergic neurotransmission in the ventral tegmental area. *Neuropharmacology* **61**, 181–188. (doi:10.1016/j.neuropharm.2011.03.026)
67. Young LJ, Lim MM, Gingrich B, Insel TR. 2001 Cellular mechanisms of social attachment. *Horm. Behav.* **40**, 133–138. (doi:10.1006/hbeh.2001.1691)
68. Caldwell HK, Albers HE 2015 Oxytocin, vasopressin, and the motivational forces that drive social behaviors. In *Behavioral neuroscience of motivation*, pp. 51–103. Berlin, Germany: Springer.
69. Gimpl G, Fahrenholz F. 2001 The oxytocin receptor system: structure, function, and regulation. *Physiol. Rev.* **81**, 629–683.
70. Dumais KM, Bredewold R, Mayer TE, Veenema AH. 2013 Sex differences in oxytocin receptor binding in forebrain regions: correlations with social interest in brain region-and sex-specific ways. *Horm. Behav.* **64**, 693–701. (doi:10.1016/j.yhbeh.2013.08.012)
71. Succu S, Sanna F, Cocco C, Melis T, Boi A, Ferri G-L, Argiolas A, Melis MR. 2008 Oxytocin induces penile erection when injected into the ventral tegmental area of male rats: role of nitric oxide and cyclic gmp. *Eur. J. Neurosci.* **28**, 813–821. (doi:10.1111/j.1460-9568.2008.06385.x)
72. Shahrokh DK, Zhang T-Y, Diorio J, Gratton A, Meaney MJ. 2010 Oxytocin–dopamine interactions mediate variations in maternal behavior in the rat. *Endocrinology* **151**, 2276–2286. (doi:10.1210/en.2009-1271)
73. Peris J, MacFadyen K, Smith JA, de Kloet AD, Wang L, Krause EG. 2016 Oxytocin receptors are expressed on dopamine and glutamate neurons in the mouse ventral tegmental area that project to nucleus accumbens and other mesolimbic targets. *J. Comp. Neurol.* **525**, 1094–1108. (doi:10.1002/cne.24116)
74. Bromberg-Martin ES, Matsumoto M, Hikosaka O. 2010 Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* **68**, 815–834. (doi:10.1016/j.neuron.2010.11.022)
75. Johnson ZV, Young LJ. 2017 Oxytocin and vasopressin neural networks: implications for social behavioral diversity and translational neuroscience. *Neurosci. Biobehav. Rev.* **76**, 87–98. (doi:10.1016/j.neubiorev.2017.01.034)
76. Baran NM, Tomaszycki ML, Adkins-Regan E. 2016 Early life manipulations of the nonapeptide system alter pair maintenance behaviors and neural activity in adult male zebra finches. *Front. Behav. Neurosci.* **10**, 58. (doi:10.3389/fnbeh.2016.00058)
77. Kao MH, Doupe AJ, Brainard MS. 2005 Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature* **433**, 638–643. (doi:10.1038/nature03127)
78. Ölveczky BP, Andalman AS, Fee MS. 2005 Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3**, e153. (doi:10.1371/journal.pbio.0030153)
79. Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, Goldberg JH. 2016 Dopamine neurons encode performance error in singing birds. *Science* **354**, 1278–1282. (doi:10.1126/science.aah6837)
80. Morrison RG, Nottebohm F. 1993 Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. *J. Neurobiol.* **24**, 1045–1064. (doi:10.1002/neu.480240805)
81. Mooney R, Hoese W, Nowicki S. 2001 Auditory representation of the vocal repertoire in a songbird with multiple song types. *Proc. Natl Acad. Sci. USA* **98**, 12778–12783. (doi:10.1073/pnas.221453298)
82. Simmonds AJ, Leech R, Iverson P, Wise RJ. 2014 The response of the anterior striatum during adult human vocal learning. *J. Neurophysiol.* **112**, 792–801. (doi:10.1152/jn.00901.2013)
83. Mandelblat-Cerf Y, Las L, Denisenko N, Fee MS. 2014 A role for descending auditory cortical projections in songbird vocal learning. *Elife* **3**, e02152. (doi:10.7554/eLife.02152)
84. Sasaki A, Sotnikova TD, Gainetdinov RR, Jarvis ED. 2006 Social context-dependent singing-regulated dopamine. *J. Neurosci.* **26**, 9010–9014. (doi:10.1523/JNEUROSCI.1335-06.2006)
85. Riters LV. 2012 The role of motivation and reward neural systems in vocal communication in songbirds. *Front. Neuroendocrinol.* **33**, 194–209. (doi:10.1016/j.yfrne.2012.04.002)
86. Gordon I, Martin C, Feldman R, Leckman JF. 2011 Oxytocin and social motivation. *Dev. Cogn. Neurosci.* **1**, 471–493. (doi:10.1016/j.dcn.2011.07.007)
87. Groppe SE, Gossen A, Rademacher L, Hahn A, Westphal L, Gründer G, Spreckelmeyer KN. 2013 Oxytocin influences processing of socially relevant cues in the ventral tegmental area of the human brain. *Biol. Psychiatry* **74**, 172–179. (doi:10.1016/j.biopsych.2012.12.023)
88. Mezey S, Csillag A. 2002 Selective striatal connections of midbrain dopaminergic nuclei in the chick (*Gallus domesticus*). *Cell Tissue Res.* **308**, 35–46. (doi:10.1007/s00441-002-0514-2)
89. Windle RJ, Kershaw YM, Shanks N, Wood SA, Lightman SL, Ingram CD. 2004 Oxytocin attenuates stress-induced c-fos MRNA expression in specific forebrain regions associated with modulation of hypothalamo–pituitary–adrenal activity. *J. Neurosci.* **24**, 2974–2982. (doi:10.1523/JNEUROSCI.3432-03.2004)

Appendix

Appendix

Chapter 1

Oxytocin and Vasopressin Receptor variants as a window onto the evolution of human prosociality *

Constantina Theofanopoulou[1,2,3,+], **Alejandro Andirkó**[1,2,+], **and Cedric Boeckx**[1,2,4,*]

[1]Section of General Linguistics, Universitat de Barcelona
[2]Universitat de Barcelona Institute for Complex Systems
[3]Laboratory of Neurogenetics of Language, Rockefeller University
[4]ICREA
[*]cedric.boeckx@ub.edu
[+]these authors contributed equally to this work

ABSTRACT

Modern humans' lifestyle strongly depends on complex social skills like empathy, tolerance and cooperation. Variation in the oxytocin receptor (*OXTR*) and the arginine-vasopressin receptors (*AVPR1A*, *AVPR1B* genes) has been widely associated with diverse facets of social cognition, but the extent to which these variants may have contributed to the evolution of human prosociality remains to be elucidated. In this study, we compared the *OXTR*, *AVPR1A* and *AVPR1B* DNA sequences of modern humans to those of our closest extinct and extant relatives, and then clustered the variants we identified based on their distribution in the species studied. This clustering, along with the functional importance retrieved for each variant and their frequency in different modern-human populations, is then used to determine if any of the *OXTR*, *AVPR1A* and *AVPR1B*-variants might have had an impact at different evolutionary stages. We report a total of 29 SNPs, associated with phenotypic effects ranging from clearly pro-social to mixed or antisocial. Regarding modern human-specific alleles that could correlate with a shift towards prosociality in modern-humans, we highlight one allele in *AVPR1A* (rs11174811), found at high frequency and linked to prosocial phenotypes in modern humans, while the ancestral allele is associated with antisocial phenotypes. We also report three sites of putatively convergent changes between modern humans and bonobos (rs237897(A), rs2228485(G) and rs1042615(A)), and note the absence of such a convergent pattern between modern humans and chimpanzees. Finally, we observe the high concentration of 'modern human specific' alleles in vasopressin receptors not paralleled in the oxytocin receptor.

1 Introduction

Oxytocin (OXT) and vasopressin (AVP) are important neurotransmitters that function through their respective receptors to regulate a diverse set of biological processes, such as pregnancy and uterine contractions, milk-ejection, copulation and orgasm, attachment between mothers and their young, bond formation, suppression of stress, thermoregulation, olfactory processing, eye-contact and recognition of familiar individuals[1]. OXT and AVP are closely related structurally and evolutionarily: they have been argued to be the product of a local duplication event that took place before the origin of vertebrates[2], and they only differ in two (of the nine) amino acids, although they display differences at a functional level[1]. Each binds to their respective receptor(s) (*OXTR* in the case of oxytocin, and *AVPR1A*, *AVPR1B*, and *AVPR2* in the case of vasopressin), but their molecular similarities allow for crosstalk in the brain and peripheral organs[3].

Variation in the genes that code for OXT and AVP receptors (*OXTR* and mainly *AVPR1A* and *AVPR1B*) have long been associated with different social behaviors[4]. Single Nucleotide Polymorphisms (SNPs) in these genes in modern humans have been claimed to be implicated in altruism, face recognition, stress levels and empathy, but also in sociocognitive disorders, such as Autism Spectrum Disorders (ASD), bipolar disorder, schizophrenia or depression[1,5]. Due to the paucity of studies on social effects of *AVPR2*, we did not include this receptor in the present study.

The role oxytocin and vasopressin play in social cognition makes them prominent candidates to test for possible social behavioral differences between hominid species (extinct and extant). In this study we examine the extent to which variation in the OXT and AVP receptors correlate with social characteristics that have already been put forth in the literature to characterize the prosocial profile of each of the species studied here (modern humans, archaic humans such as Neanderthals and Denisovans, bonobos and chimpanzees). 'Prosociality' is a broad term that encompasses intraspecies empathy, social tolerance, cooperation

and altruism. While our closest living relatives, the chimpanzees (*Pan Troglodytes*) and the bonobos (*Pan Paniscus*), live in highly organized social groups as well, present-day humans' social networks are larger and denser, powered by a complex social cognitive machinery[6]. Modern humans are characterized by great intrasocial compassion, are motivated by concern about the welfare of out-group individuals, and display a clear tendency to act in concert, to the extent that *Homo Sapiens* has been labeled as 'ultra-social'[7]. This trait is of special relevance, as it has been argued to underlie other singular traits of humans, such as their enhanced verbal communicative skills[6–10].

The sequencing of two Neanderthal genomes from Altai (Siberia)[11] and Vindija (Croatia)[12] and a Denisovan from Altai[13] has made available genomic data to provide new insights into the discussion of the evolution of social cognition, complementing the archaeological evidence. Today, various hypotheses[6,10,14,15] still offer different explanations and timelines for the emergence of prosociality, ranging from the *Pan-Homo* split to later stages of human evolution, such as the split between Neanderthals and Denisovans on the one hand, and Modern Humans on the other. The critical effect of OXT and AVP on pair-bonding has led some of the authors of the aforementioned theories, most prominently,[14], to ascribe to them a key role in the emergence of human social behavior, while others have challenged the centrality of OXT and AVP in this shift in favor of other hormones, such as β-endorphines and dopamine[10,16]. By examining the evolutionary variation in human OXT and AVP receptors, we aim to shed light onto the timing of the transition towards the current status of human prosociality, as well as determine more clearly the specific role that OXT and AVP could have played in this regard.

As of now, none of the studies searching for fixed changes between modern and archaic humans (Neanderthals and Denisovans) have identified changes on the genes coding for the OXT and AVP ligands and receptors[11,17]. The only study[17] systematically exploring non-synonymous changes at high frequency in modern humans for which archaic humans carry the ancestral state found that *AVPR1B* is in the top 5% of the genes enriched for high frequency-changes in modern humans (controlling for gene length).

For this reason, in this study we investigated the variants that differ in modern and archaic humans on the *OXTR*, *AVPR1A* and *AVPR1B* genes, focusing on those that are polymorphic in modern humans and that have been associated with specific behavioral correlates in the literature, using also allele-frequency data from modern humans of different ethnic backgrounds. In order to infer the ancestral state (allele) of these sites we used primate species' sequences (rhesus macaque, chimpanzee, bonobo). We also took into account variation data (Single Nucleotide Variants: SNVs) from multiple chimpanzee and bonobo individuals. We identify various changes in the analyzed genes which we clustered in different evolutionary stages, based on their distribution (presence or absence) in the different species/populations studied (e.g. Homo-specific, modern human-specific, Altai Neanderthal-specific). These changes have been reported in the literature to affect gene expression, brain regions such as the mesolimbic reward system, and behavioral phenotypes. A fair amount of those polymorphic sites also confer risk of sociocognitive disorders, like Autism Spectrum Disorder (ASD). Finally, we discuss how the information we have gathered bears on several hypotheses concerning the evolution of human prosociality, including the neurochemical hypothesis[10], the social-brain hypothesis[6] and the self-domestication hypothesis[14,15].

2 Results

The DNA sequence-alignment we performed gave rise to a list of SNPs that we ordered in clusters (shown in Tables 1-2), based on their distribution in the sequences studied, with the major distinction being SNPs present only in modern humans (MHS: modern human-specific) *vs.* SNPs shared between modern humans and one or more archaics, and those distiguishing *Homo* from *Pan*. In this section we present the SNPs we identified, along with their potential functional relevance, based on data mining as well as our independent analysis (SNAP2 test). We discuss the results following their distribution pattern: from total overlap (alleles found in all the species considered) to no overlap at all (e.g., alleles found exclusively in modern populations, or MHS), and summarize the key information in Table 1 (for the oxytocin receptor) and Table 2 and 3 (for vasopressin receptors). Figures 1-2 provide graphic summaries of the main results. Frequencies of the relevant alleles in modern human populations retrieved from the sources consulted (see Methods) are provided in Supplementary Tables 1-2. A series of archaic human-specific variants were also identified and are reported in Supplementary table 3. Just one of them (rs199856198, G/A) was found to be an extremely rare allele in modern humans ($<$0.002). Rs199856198 is a missense variant in exon four of *OXTR* that changes Threonine for Methionine at the 360th position. While its effects have not been investigated, the SNAP2 test we performed gave a predicted 63% for a possible effect on the phenotype.

Only four alleles discussed here are not shared by the three non-human primates we used: rs237897(A) and rs2228485(G) are shared between modern humans and only bonobos; and rs11131149 (A) is found also in rhesus macaques (*Macaca mulatta*), but not in chimpanzees or bonobos.

2.1 Oxytocin receptor

The intronic variants rs11131149(A), rs59190448(G) and rs13316193(C), the 3'-UTR variant rs9872310(G) and the missense SNP rs4686302(T) are found in both present-day populations and the three ancient human sequences used in this study.

Rs11131149(A), already attested in macaques, has been reported to have the reverse effect of the G allele, which is found in chimpanzees and bonobos and correlates with higher social performance (empathy, joint attention, cooperation and self-recognition) in 18 month-children[18]. Interaction between the G allele and maternal cognitive sensitivity accounted for a 26% of variability in a Theory of Mind scale in 4.5 year old-children[18]. Rs11131149(G) is also part of a haplotype related to depressive temperament[19].

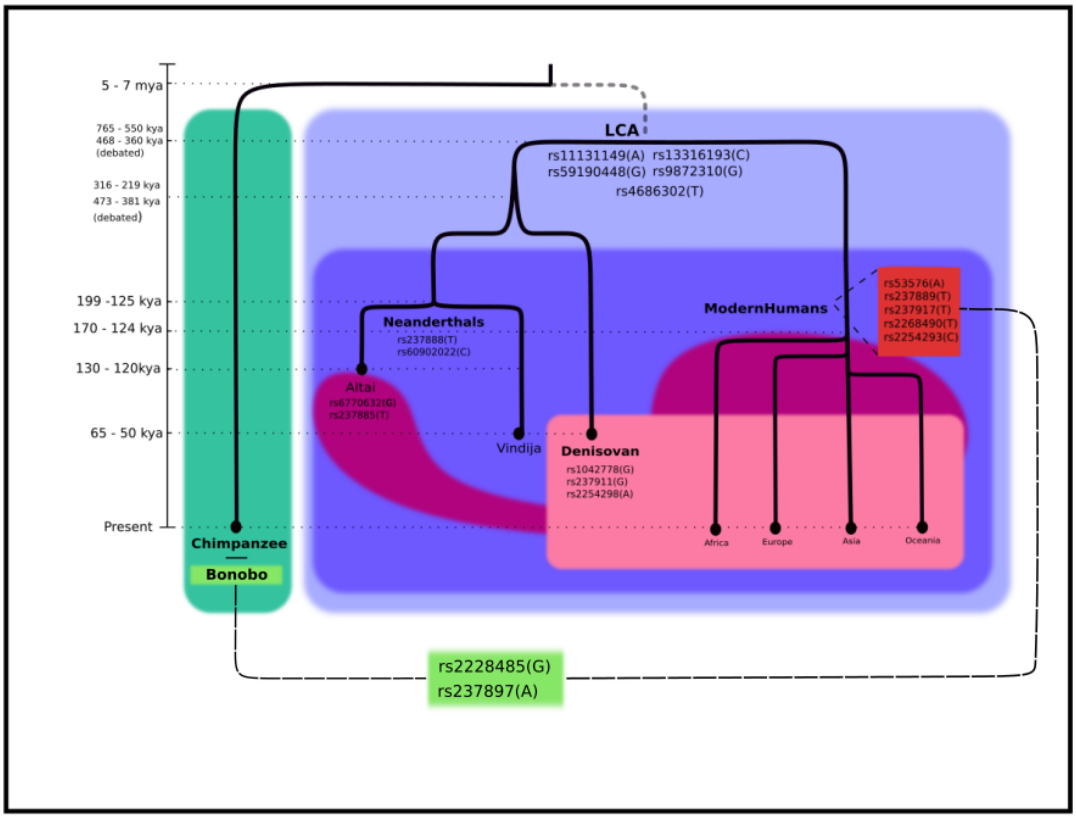


Figure 1. Evolutionary distribution of *OXTR* alleles. The alleles displayed are the non-ancestral ones. LCA = Last Common Ancestor.

Rs59190448(G) has been argued to show signs of positive selection in present-day humans[20]. The only known endophenotype associated with it is increased risk of anxiety, stress and depression in early life[21]. Rs13316193(C) has been related to empathy[22] and high cooperation and comforting skills[23], but also to late onset of Obsessive Compulsive Disorder[24], poorer social skills[25] and significant association with Attention Deficit/Hyperactivity Disorder (ADHD) on the Social Communication Questionnaire[26]; rs13316193(T) is part of a haplotype linked to ASD[27], depressive mood[19], and poorer empathic communication in relationships[28]. The T allele also affects *OXTR* total gene expression in the brain[29]. Rs9872310(G) has been implicated in altruism and ASD in different studies[27,30], but its specific functionality has not been investigated further. The rs4686302(T) allele benefits perspective taking[22] and social connectedness (in men)[31] compared to the C allele, while ADHD T-carriers performed significantly worse on the face emotion recognition task than C-carriers[31].

OXTR alleles rs237888(T) and rs60902022(C), both intronic, are found in both (Altai and Vindija) Neanderthal sequences but are absent in the Denisovan sequence. The ancestral allele rs237888(C) has been associated with daily life-skills score in the Vineland Adaptive Behavior Scales (VABS) test in ASD patients, as well as with IQ measurements[27], and the T allele has been linked to greater impairment in ASD[32]. Rs237888 is part of a haplotype related to altruism in the Dictator Game[30](an experimental economics paradigm where participants have to assign amounts of money to different individuals) and it has been also been associated with DNA methylation of specific CpG sites (cg25140571 and cg00247334) that are linked to abuse and

psychiatric symptoms[33]. Rs60902022(C) has been claimed to affect gene expression and transcription factor binding by linkage disequilibrium (LD) with other *OXTR* variants[34].

In addition to these SNPs, we identified in the Altai Neanderthal *OXTR* sequence two present-day human alleles not found in the Vindija and Denisova sequences: rs6770632(C) and rs237885(T). The 3'UTR rs6770632(G) has been associated with VABS scores[27] and persistent, extreme aggression, with the C and T alleles affecting male and female children, respectively[35]. rs237885(T) has been associated with callous/unemotional traits[35], ASD[36], schizophrenia diagnosis[37] and higher risk of aggression[38], while the G allele is linked to altruistic allocations in the Dictator Game[30].

Modern human alleles found in the Denisovan individual but not in Altai or Vindija Neanderthals are rs1042778(G), rs2254298(A) and rs237911(G). The T allele of rs1042778 has been associated with lower levels of OXT in plasma, diminished parental care (parent-child gaze and touch)[39] and panic/aggressive behaviors[40], while the G allele has been linked to ASD[27,40,41] and to aggression in males[35]. However, this latter association is at odds with other findings concerning the G allele reporting a significant correlation with prosocial fund allocations in the Dictator Game setting[30] (although[42] failed to replicate the result) and higher scores in altruistic and comforting behaviors[23]. According to[43], T-allele carriers are likely to recover from the effects of low maternal emotional warmth and acceptance, whereas G-carriers do not show such a pattern. But based on another study[44], it was G allele-carriers who experienced gains in daily positive emotions from loving-kindness training, whereas individuals with the T allele did not. Additionally, it has been suggested that rs1042778(G) influences *OXTR* transcription and translation processes, as well as *OXTR* gene expression in the amygdala[29,30].

Rs2254298(A) and rs237911(A) are overtransmitted in ASD patients[36,45], a result confirmed in a meta-analysis that included eleven cohorts[46]. Such effects seem to depend on ethnicity, as[45] and[36] used Chinese and Japanese samples, while a study using a Caucasian sample found rs2254298(G) to be the variant associated with ASD[27,47]. rs2254298(G) has been associated with lower communication scores in romantic relationships[28], variation in empathy scores[22], methylation at cg11589699 (a site linked to depression and anxiety level increase)[48], less sensitive parenting and lower plasma OXT levels[39], but also with higher values of positive affect and lower scores in depressive temperament in a Japanese sample[19].

Rs2254298(A) carriers performed better in self-reported empathy[37] and empathy for pain in particular[49], parenting[50] and in attachment security tests (in a non-Caucasian children sample)[51], while A-ADHD-carriers displayed fewer social deficits[26]. On the contrary, A-ASD-carriers presented more social deficits[26] and lower serum OXT-levels[52]. This allele has also been related to prosopagnosia[53], high levels of physical aggression and hostility[54] and low emotion recognition and resilience skills[55]. G-carriers showed higher levels of retrospective self-report of inhibition and adult separation anxiety[56] and, compared to A-carriers, are more vulnerable to antisocial behavior if they experience maltreatment[57]. This SNP also has interesting anatomical associations: the A allele was associated with larger amygdalar volume in healthy Asian adults[58,59], a phenotype typically identified in the early stages of autism[58], and which correlated with heightened amygdala response during two functional magnetic resonance imaging (fMRI) tasks that involved viewing socially-relevant face stimuli[59]. However, this association was not replicated though in a healthy Caucasian sample[60]. Gender might be playing a role in these associations, since A-female carriers showed smaller left amygdala volume, while it was G-male-carriers that showed smaller left amygdala, which was also negatively associated with attitudinal trust[61].

Finally, the intronic alleles rs2268490(T), rs2268493(C), rs237889(T), rs237917(T), and rs53576(A) were only present in modern human populations. In addition, the intronic rs237897(A) and the synonymous variant rs2228485(G) are only attested in modern human populations and in bonobos, and are thus putative instances of convergent evolution.

The archaic rs2268490(C) allele positively affects the amount of funds altruistically given in the Dictator Game setting[30] and might provoke vocal alterations under stress. Carriers of the MHS C allele displayed more stress-related vocal symptoms (dysphonia, muscle tension, frequency changes) and higher cortisol levels[62]. The MHS allele rs2258493(T) has been linked to ASD subphenotypes[63] and diagnosis[41,64,65], negative scores in social performance, perception and mentalizing tasks in schizophrenia patients[66], ADHD patients[67] and depressive temperament (as part of a haplotype block)[19]. Carriers of this allele also showed reduced mesolimbic reward system activation, a result that might point towards the neurobiological basis of the aforementioned phenotypic effects of this SNP[68].

Rs237889(T) has been associated with ASD, both as part of a deleterious haplotype[27] and independently[69], as well as with differences in moral judgment; carriers of the archaic C allele were more prone to give utilitarian answers in dilemmas[70]. Rs237897(A) is part of a haplotype related to ASD[27], altruism in males[30], lower self-reported betrayal levels[71], continuous social connectedness[72], and Theory of Mind[18]. Alleles of rs53576 have been reported in several studies: the G allele has been reported to be implicated in Bullimia Nervosa[73], but also in diminished stress after social support[74], adult separation anxiety[75], oxytocin sensitivity in social cooperation settings (increased in males, decreased in females)[76], overall weak social cognition skills in ADHD patients[25] and facial recognition deficits[53]. MHS rs53576(A) might be involved in ASD[27,45], higher empathic performance[77,78] and social connectedness in women[31], but also lower psychological resources such as self-esteem, optimism and emotional mastery[79]. Though the literature on rs53576 doesn't provide unequivocal results, there seems to be consensus on this SNP being dependent on environmental factors: the G allele appears to affect social sensitivity; adverse life conditions

can lead to negative (non-prosocial) behavior in G carriers, but the opposite effect has also been reported[80]. Rs237917(T) is related to emotion recognition[81]. SNP rs2228485(A) is part of a haplotype related to loneliness[82] and overtransmitted in ASD[45]. Carriers of the G allele were more prone to give incorrect answers when required to identify negative emotions in male face images[83].

Table 1. Allele distribution on the *OXTR*-SNPs in the species studied. Al: allele, Pop.: population, Ancestral: also found in Macaques, Bonobos, Chimpanzees. Ancestral: also found in bonobos. Homo: also found in Archaic (Neanderthals, Denisovans) and Modern Humans. Abbreviations: MH: Modern Humans, MHS: Modern Human-Specific, Neand: Neanderthals, Den: Denisovan, B: Bonobos, TFB: Transcription Factor Binding, LD: Linkage Disequilibrium.

SNP	Type	Al.	Pop.	Effect	Trial sample	Other remarks
rs2228485	Exonic	A	Ancestral	Loneliness[82]	285	Synonymous. A/G in Bonobos, A in Chimpanzees
				ASD[45]	195 (Chinese)	
		G	MHS (also present in B.)	Emotion recognition[83]		
rs237897	Intronic	G	Ancestral			G/A in Bonobos, G in Chimpanzees
		A	MHS (also present in B.)	ASD[27]	152	
				Altruism[30]	203	
				Lower self-reported betrayal levels[71]	165	
				Social connectedness[72]	11.000	
				Theory of Mind[18]	301	
rs11131149	Intronic	G	Ancestral	Theory of Mind, higher levels of social cognition[18]	350 children	
				Depressive mood[19]	493 (Japanese)	
		A	Homo	Lower levels of social cognition[18]	350 children	Present in macaque
rs59190448	Intronic	A	Ancestral			
		G	Homo	Anxiety, stress and depression risk[21]	653	Positive selection[20]
rs13316193	Intronic	T	Ancestral	ASD[27]	152	
				Depressive mood[19]	493 (Japanese)	Affects *OXTR* total expression in the brain[29]
				Poor empathic communication[28]	120	
		C	Homo	Empathy[22]	101(Chinese)	
				Poor social skills[25]	112	
				Greater cooperation and comforting[23]	422 (Chinese males)	
				ADHD[24]	276 ADHD patients	
				Face emotion recognition[26]	151 children with ADHD	
rs9872310	3' UTR	A	Ancestral			
		G	Homo	Altruism[30]	203	
				ASD[27]	152	
rs4686302	Exonic	C	Ancestral			Missense
		T	Homo	Better perspective taking skills[22]	101(Chinese)	
				Face emotion recognition[84]	151 children with ADHD	

				Social connectedness in men, opposite in women[31]	Over 11000 individuals	
rs237888	Intronic	C	Ancestral	IQ and VABS scores[27]	152	
				Altruism[30]	203	
		T	MH+Neand	Greater impairment of ASD[32]	1002 ASD patients	
				Methylation of CpG sites linked to abuse and psychiatric symptoms[33]	393 African American adults	
rs60902022	Intronic	T	Ancestral			May affect TFB through LD[34]
		C	MH+Neand			
rs6770632	3' UTR	A	Ancestral	Aggression[35]	160 children	
		G	MH+Altai	VABS scores[27]	152	
rs237885	Intronic	G	Ancestral	Altruism[30]	203	
		T	MH+Altai	ASD[36]	282 (Japanese)	
				Schizophrenia[37]	145	
				Callous/unemotional traits[35]	160 children	
				Higher risk of aggression[38]	488 cases, 488 control (Chinese)	
rs1042778	3' UTR	T	Ancestral	Lower levels of OXT in plasma, diminished parental care[39]	352	Affects *OXTR* transcription and translation processes, amygdalar expression[29,30]
				Panic and aggressive behaviors[40]		
				Recovery from low maternal emotional warmth[43]	2341	
		G	MH+Den	ASD[27,40,41]	152, 2333, 209	
				Aggression[35]	160 children	
				Prosocial fund allocations in the Dictator Game[30]	203	
				Might lower transcription levels of *OXTR*[40]		
				Altruism, comforting behavior[23]	422 Chinese males	
				Positive emotions after training[44]	122	
rs237911	5' UTR	A	Ancestral	ASD[36,45,46]	195 (Chinese), 282 (Japanese), 3941	A/C in macaque
		G	MH+Den			
rs2254298	Intronic	G	Ancestral	Lower communication[28]	120	
				Variation in empathy[22]	101 (Chinese)	
				Methylation at cg11589699 (increased depression and anxiety)[48]	393 (African American)	
				Less sensitive parenting and lower plasma OXT[39]	352	
				Higher positive affect[37]	352	
				Lower scores in depressive temperament[19]	493 (Japanese)	
				Higher levels of Retrospective Self-Report of Inhibition and Adult Separation Anxiety[56]	93 patients	
				Smaller left amygdala[61]	211 (men), 199 (women)	
		A	MH+Den	ASD[36,45,46]	195 (Chinese), 282 (Japanese), 3941	

				Lower levels of emotion recognition and resilience scores[55]	264 (Korean)	
				Increased amygdala volume[59]	55	
				Fewer social deficits in an ADHD sample more social deficits in an ASD sample[26]	341 (ASD patients), 276 (ADHD patients)	
				Lower serum OT in ASD patients[52]	55 (ASD patients), 110 (controls)	
				Positive parenting behavior, physically controlling behavior[50]	157 mothers	
				Reponsive to the impact of adversity[49]	302	
				High levels of physical aggresion[54]	197 (Chinese) adolescents	
				Vulnerability for antisocial behavior after maltreatment	1591	
rs53576	Intronic	G	Ancestral	Bullimia Nerviosa[73]	262 (Korean)	
				Diminished stress[74]	176 (77 Caucasian, 99 non-Caucasian)	
				Separation anxiety[75]	185	
				Oxytocin sensitivity in social cooperation settings (increased in males, decreased in females)[76]	204	
				Weak social cognition in ADHD[25]	112	
				Facial recognition deficits[53]	18 (Italian), 6 (German)	
		A	MHS	ASD[27,45]	195 (Chinese), 152	
				Empathy[77,78]	50, 192 (multiple ethnicities)	
				Lower psychological resources[79]	344	
				Social connectedness (women)[31]	Over 11000	
rs2268490	Intronic	C	Ancestral	Altruism[30]	203	
				Vocal alterations under stress[62]	657 (Finnish twins)	
		T	MHS	Stress-related vocal symptoms and higher cortisol levels[62]	657 (Finnish twins)	
rs2268493	Intronic	T	Ancestral	ASD[41,63–65]	417 (multiple ethnicities), 530 (Caucasian), 527, 2.333	Reduced mesolimbic reward system activation[68]
				Negative scores in social tasks in schizophrenia[66]	74	
				ADHD[67]	99	
				Depressive temperament[19]	493 (Japanese)	
		C	MHS			
rs237917	Intronic	C	Ancestral			
		T	MHS	Emotion recognition[81]	207 (Central European)	
rs237889	Intronic	C	Ancestral	Utilitarian answers in dilemmas[70]	228, 322	
		T	MHS			
				ASD[27]	152	

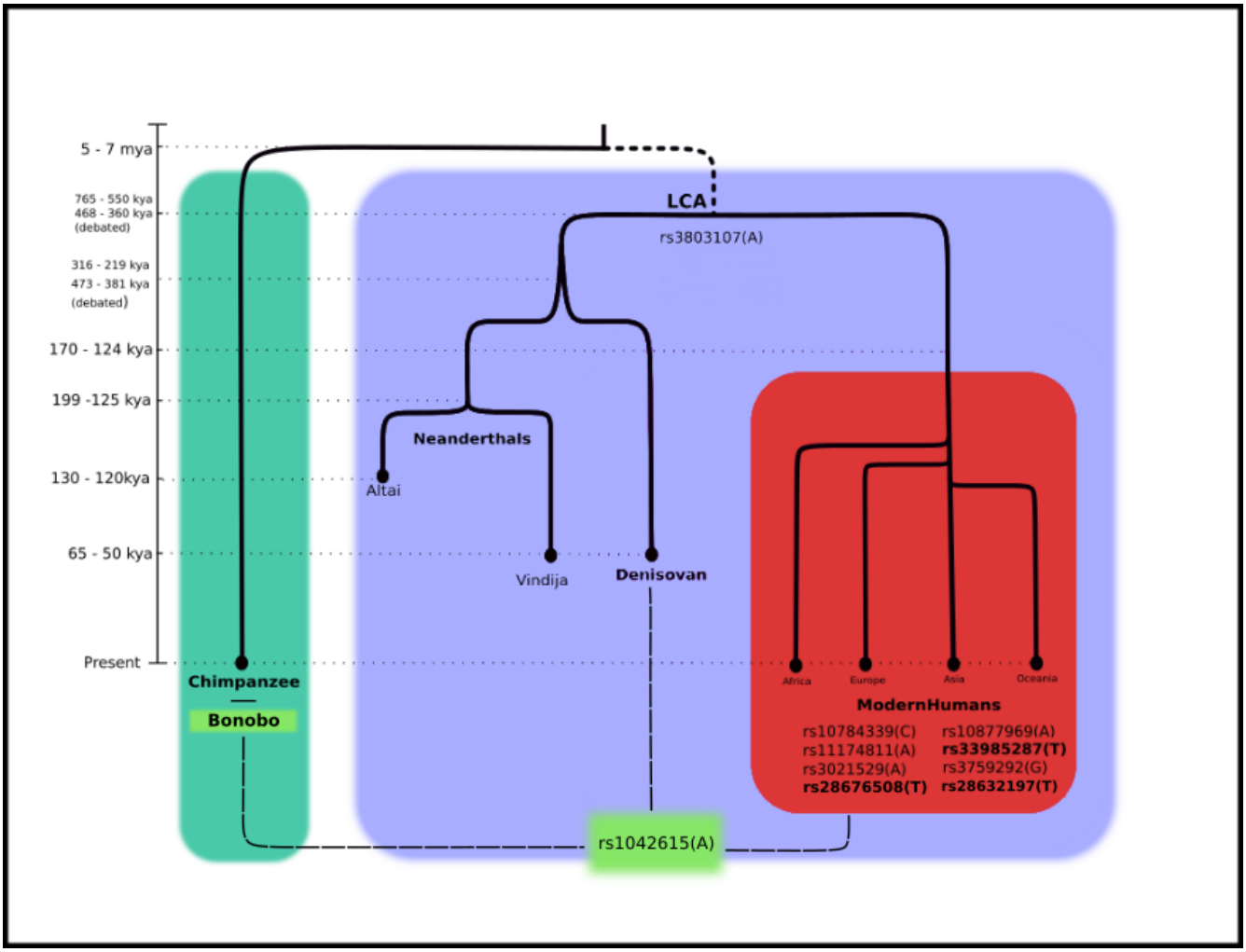


Figure 2. Evolutionary distribution of *AVPR1A* (regular) and *AVPR1B* (bold) alleles. The alleles displayed are the non-ancestral ones. LCA = Last Common Ancestor.

2.2 Vasopressin receptors

The distribution of vasopressin receptors is somewhat less complex than that of the oxytocin receptor. For example, we could not identify any sites that are both polymorphic in modern humans and different within the two neanderthals included in our study.

Only one modern human allele of *AVPR1A* was identified in both the Neanderthal and Denisovan genomes: rs3803107(A). Rs3803107(A) (3'-UTR) has been studied in relation to ASD in an Irish sample, but this correlation did not reach the level of significance[85]. Rs1042615(A), a synonymous variant of *AVPR1A*, also showed association with ASD in present-day humans[86] and often occurring vocal symptoms during stress[62], but in the ancient DNA sample it was only found in the Denisovan individual. Rs1042615(A) is the third site in this study that is also found in bonobos, constituting another potential convergent site.

The ancestral G allele of the 3'-UTR variant rs10784339 has been associated with stress reactivity and substance addiction risk[87,88], while the function of the MHS C allele is unknown. The ancestral C allele of rs11174811 (3'UTR) is related to substance addiction risk[87,88], but also to higher anxiety levels[89] and aggression[35]. The MHS variant disrupts a microRNA binding site, increasing the expression levels of *AVPR1A* and possibly affecting the anxiety relief consequences of vasopressin in anxious situations[88].

The ancestral G allele of rs3021529 may also be under balancing selection and affect the regulation of the gene[20], and has been linked to addiction[90]. The ancestral A allele of rs3759292 was found to be under directional selection[20], but without any reported functional implications. The MHS G allele has been linked to heroin addiction[91] and also to ASD[92]. Other alleles have been also studied in the context of social behavior and related disorders, especially ASD, such as the MHS rs10877969(A)

(intron variant)[92,93]. Concerning *AVPR1B*, rs28676508(T) has been claimed to be involved in child onset aggression[94]. The missense (arginine to histidine, position 364) variant rs28632197(T) has been associated with ASD diagnosis[63] and panic disorder[95]. Finally, the G allele of rs33985287 protects against depressive moods in female children[96].

Table 2. Allele distribution on the *AVPR1A*-SNPs in the species studied. Ancestral: Also found in Macaques, Bonobos, Chimpanzees. Ancestral: also found in bonobos. Homo: Found in Archaic (Neanderthals, Denisovans) and Modern Humans. Abbreviations: MH: Modern Humans, MHS: Modern Human- Specific, Neand: Neanderthals, Den: Denisovan. B: Bonobos.

SNP	Type	Alleles	Pop.	Effect	Trial sample	Other remarks
rs1042615	Exonic	G	Ancestral			Missense. G/A in Bonobos, G in Chimpanzees
		A	MHS+Den (also present in B.)	ASD[86]	205 (Finnish)	
rs3803107	3' UTR	G	Ancestral			
		A	Homo			
rs10784339	3' UTR	G	Ancestral	Stress reactivity and substance addiction risk[87,88]	852, 2231	
		C	MHS			
rs11174811	3' UTR	C	Ancestral	Substance addiction risk[87,88]	852, 2231	Possibly under balancing selection[20]
				Higher anxiety levels[89]	1090 (German)	Increases expression of *AVPR1A*
				Aggression[35]	160 children	
		A	MHS			
rs3021529	Intronic	G	Ancestral	Addiction[90]	1.050	Possibly under balancing selection[20]
		A	MHS			
rs10877969	Intronic	C	Ancestral			Except macaque (G)
		A	MHS	ASD[92,93]	151 Korean trios, 633	
rs3759292	Intronic	A	Ancestral			Positive selection[20]
		G	MHS			

3 Discussion

This study reports a total of 29 SNPs, 19 for *OXTR*, and 10 for *AVPR1A* and *AVPR1B*. Of these, 5 and 8 variants, respectively, are MHS, which means 80% of the total of mutations in the case of AVP receptor genes. In addition, 3 variants (2 for *OXTR*, 1 for *AVPR*) are putative convergent sites between modern humans and bonobos. Only some of these SNPs (rs59190448, rs3021529, rs11174811, and rs3759292) have been previously claimed to be under selection in modern humans. There is evidence linking some of the SNPs identified here with prosocial behaviors (rs237917, rs2268490, rs237885 [section 2.1]; rs11174811 and rs33985287 [section 2.2]). The rest of the SNPs are either neutral, give mixed results, or confer risk of some social behavior-disorder, mainly ASD. Some of the limitations of this study listed at the end of this article may contribute to these results.

The clearest pattern we detect concerns AVP receptors, specifically, *AVPR1A*. 3 of the 5 MHS alleles (on rs11174811, rs3021529, rs3759292, all of which have been associated with signals of selection) occur at very high frequencies in the global population (Table S2). Of these, the A allele of rs11174811 shows the clearest change towards prosocial effects (the archaic C allele is associated with negative phenotypes). Such a change from a more ancient allele linked to negative effects to a

Table 3. Allele distribution on the *AVPR1B*-SNPs in the species studied. Ancestral: Also found in Macaques, Bonobos, Chimpanzees. Ancestral: also found in bonobos. Ancestral(CB) also found in Chimpanzees and Bonobos. Homo: Found in Archaic (Neanderthals, Denisovans) and Modern Humans. Abbreviations: MH: Modern Humans, MHS: Modern Human- Specific, Neand: Neanderthals, Den: Denisovan.

SNP	Type	Alleles	Pop.	Effect	Trial sample	Other remarks
rs28676508	Exonic	C	Ancestral			Synonymous. C/G in Bonobo, Chimpanzee
		T	MHS	Child onset aggression[94]	177	
rs28632197	Exonic	C	Ancestral			Missense. C/G in Bonobo, Chimpanzee
		T	MHS	ASD[63] Panic disorder[95]	207 186 (German)	
rs33985287	3' UTR	C	Ancestral	Protects against depressive moods[96]	464 (children)	C/G in Bonobo, Chimpanzee[97]
		T	MHS			

MHS allele linked to positive effects occur five times in our data: three times for *AVPR1A* (rs10784339 G>C, rs11174811 C>A and rs3021529 G>A), and two for *OXTR* (rs2268493 T>C and rs237917 C>T). But of these changes, only the *AVPR1A* rs11174811(A) reaches near-fixation in modern human populations. Comparative work on chimpanzees and bonobos[98,99] has highlighted the relevance of OXT and AVP receptors, especially *AVPR1A*, to capture differences in social cognition. Our analysis points in the same direction for archaic vs. modern humans.

Our analysis of *OXTR* yields more mixed results. Only one MHS mutations (on rs237917) is associated with positive effects. As a matter of fact, some alleles associated with negative phenotypes (rs59190448, rs237911) occur at high frequences in several populations (Table S1). Other alleles that occur at high frequencies in most modern populations (rs9872310, rs4686302, rs2268493, rs33985287) lack clear phenotypical effects. While the change on rs4686302 could have boosted prosociality, our SNAP2 test showed that this site is most likely of no functional importance (82% accuracy).

Taken on its own, the evolutionary distribution of *OXTR* alleles could be taken to lend some support to hypotheses that argue for early changes in our lineage associated with prosocial behavior, unlike the changes on *AVPR1A* and *AVPR1B* that appear to be largely clustered in MHS. It is certainly compatible with hypotheses like the neurochemical hypothesis put forth in[10], or the series of pro-social steps defended in[6]. Although these accounts stress the role of other hormones in early changes in hominins (dopamine in the case of[10] and β-endorphines for[6]), all of these hormones (especially oxytocin and dopamine) are known to interact and reinforce each other's effects[16,100], so it could be that the early changes in *OXTR* identified here formed part of a broader set of changes, early in our clade, that set the stage for our prosocial profile.

Still, our results, especially those concerning the AVP receptors, also point to a distinct MHS social profile, which meshes well with the predictions of another working hypothesis that tries to account for modern humans' prosociality, the 'self-domestication hypothesis'. Advocates of this hypothesis[14,15,101], build their case on certain physiological and behavioral traits that modern humans share with domesticated animals to argue for a significant turning point exclusive of *Homo sapiens* on the prosocial continuum. Although he does not endorse the logic of self-domestication,[6] also recognizes a special transition corresponding to the emergence of our species. Among these traits, digit ratio—a measure of prenatal androgen exposure[102]—suggests that Neanderthals had higher prenatal androgen exposure than modern humans[103]. Interestingly, one study reports that the association between digit ratio and cognitive empathy is contingent on one of the *OXTR* SNPs (rs53576) we mentioned in the Results, showing a three-way association between testosterone, oxytocin and empathy[104]. In the context of the self-domestication hypothesis, it is worth pointing out that both oxytocin and vasopressin receptors have been found to be under relaxed selective constraint in domesticated species[105], and have been claimed to facilitate domestication[106].

Our results could be used as a springboard for other studies delving into the differences in prosociality between bonobos and chimpanzees, as well as for those studies looking into evidence for convergent evolution in bonobos and modern humans in an attempt to explain their similarities in terms of prosociality[107,108]. We found three alleles that bonobos and modern humans share (rs237897(A), rs2228485(G) and rs1042615(A)), while we did not find any for modern humans and chimpanzees. Of these only rs1042615(A) is a missense mutation, while rs2228485(G) is synonymous and rs237897(A) an intronic variant. Even though missense mutations tend to attract more scientific interest, there is accumulating evidence that synonymous SNPs can affect splicing or mRNA stability, thereby altering gene products[109]. The association studies on these sites give mixed

results, so it would be interesting to pursue these sites' functionality further in a larger bonobo sample.

Among the Neanderthals we found that only the Altai carried two present-day alleles which have been associated with antisocial behavior, such as ASD, schizophrenia, (female) aggression (section 2.1) and *OXTR* mRNA expression in the brain[25]. If it is the case that these SNPs were frequent and not a fabric of the small sample of ancient human DNA currently available, it could mean that within the general Neanderthal population, Altai Neanderthals might have been less social than their conspecifics of other populations. A less prosocial attitude would be consistent with the high inbreeding rates found in the genome of the Altai Neanderthal[11]. According to[110], Neanderthals were deeply subdivided into small population groups with scarce contact between them, which may have given them a social profile distinct from *Homo sapiens*.

SNPs present only in present-day humans and the Denisovan individual are of special interest considering the lack of archaeological information on Denisovans. According to paleogenomic studies, the rate of inbreeding of the sequenced individual is high, suggesting a very low population size alongside a two-fold increase of *H. Sapiens* competitor population size[13]. Some of these differences might be modulated by *OXTR* variation (rs1042778 and rs1042615 increase ASD-risk, while the first one also affects altruism positively (sections 2.1, 2.2)).

We acknowledge that there are limitations to this study. First, there are vastly more genomes currently available for the modern human population. While this may tip the balance towards modern human specificity in our study, the contrasting patterns obtained for oxytocin and vasopressin receptors suggest that our results cannot be fully reduced to the number of genomes available. Second, we have assumed that the SNPs studied would have the same (if any) effect on archaic humans or great apes, while their functionality has only been studied in modern humans. Since we are dealing with different genomic backgrounds, our interpretation remains tentative, although it is broadly compatible with information based on the fossil record and paleogenomic evidence (like inbreeding rates) or with behavioral differences between chimpanzees and bonobos. Also different plasticity windows have been hypothesized to play a role in susceptibility to both positive or negative influences[111]. Thus, it could be that the different ontogenetic trajectories that have been hypothesized for modern humans and Neanderthals[112] based on fossil evidence shaped a different susceptibility profile for them. Third, we have assumed that the ancient genomes that have been sequenced were representative of the general archaic population, something that might not be the case. Fourth, the allele-distribution data (Tables S1 and S2) we found in the literature for different modern human populations come from studies that have used different sample sizes, thus it might be that the high distribution of an allele is in reality a false positive. For this reason, we have limited our analysis of these tables to the Discussion. Fifth, all the sites that we considered here and labeled polymorphic in chimpanzees and bonobos (rs2228485, rs1042615, rs28676508, rs28632197, rs33985287) were in fact present with a 100% frequency in all the individuals of the SNV-data we used, but they differed from the allele present in the reference genomes. For this reason, in order to infer the ancestral state, we also made use of the gorilla and the orangutan genomes (apart from the macaque), which in all these sites showed the same variants as in the chimpanzee and bonobo reference genomes. Future research should use larger population samples to figure out the state of these sites. Sixth, our study may suffer from a publication bias where alleles with negative effects are overrepresented because of their clinical relevance. Finally, it could be said that our study favors oxytocin and vasopressin instead of other hormones, such as β-endorphines, cortisol, dopamine and testosterone, that have also been claimed to have been crucial in the evolution of our prosociality. While we have conveyed that there is enough theoretical ground to choose OXT and AVP for this study, we have also acknowledged that the role of oxytocin and vasopressin in prosociality depends on its interactions with other hormones that regulate social behavior.

4 Methods

We retrieved the *OXTR*, *AVPR1A* and *AVPR1B* DNA sequences from the following sources: the publicly available genomes of two Neanderthals and a Denisovan[11–13], seven high-coverage present-day human genomes (San(HGDP01036), Mbuti(HGDP00982), Karitiana(HGDP01015), Yoruba(HGDP00936), Dinka(DNK07), French(HGDP00533) and Han(HGDP00775) genomes, originally sequenced for[12]), 1000 Genomes project[113], manipulated through the Ensembl[114], the chimpanzee (*Pan Troglodytes*) genome (CHIMP2.1.4 version), the bonobo (*Pan Paniscus*) genome (PANPAN1.1, Max-Planck Institute for Evolutionary Anthropology version) and the rhesus macaque (*Macaca Mulatta*) genome publicly provided by Ensembl[114]. We also used Single Nucleotide Variant (SNV)-data found in[97] for 13 bonobos (Pan paniscus) and 25 chimpanzees covering from west to east Africa (10 *Pan troglodytes ellioti*, 6 *Pan troglodytes schweinfurthii*, 4 *Pan troglodytes troglodytes*, 4 *Pan troglodytes verus*, and 1 chimpanzee hybrid).

Alignments were performed with the following tools: the built-in Ensembl tool[114], the Max Planck for Evolutionary Anthropology Ancient Genome Browser (https://bioinf.eva.mpg.de/jbrowse/), Aliview[115], Decipher for R[116], Bedtools, MUSCLE[117] and MView[118]. We used all the genomic sequence of the genes we aligned, as provided in the standard layout of the files of the genomic sequences in the Ensemble database, namely with 600 bp upstream and downstream. We defined the genomic sequences in the same way when we extracted the gene sequences from the archaic genomes. We found no gaps in the gene sequences we studied in archaic humans (Altai and Vindija Neanderthals and Denisovans). We used the Integrative Genomics Viewer (IGV)[119] to search for the relevant SNP-positions in the bonobo and chimpanzee SNV-data.

We first aligned the modern human gene sequences of *OXTR*, *AVPR1A* and *AVPR1B* against each archaic human gene sequence and of the differences we found, we focused on those which are polymorphic in modern humans. We then aligned the modern human sequences *OXTR*, *AVPR1A* and *AVPR1B* against the chimpanzee, bonobo and macaque sequences in order to infer the ancestral state of previously identified sites (Table S5). The SNV-data from bonobos and chimpanzees were aligned to the *hg38*; we searched ad hoc for the locations of the SNPs of interest to account for variation in these sites. All alleles we studied were present with a 100% frequency in the SNV-data. When the allele found in the SNV-data was different from the allele present in the reference genomes (as in rs2228485, rs1042615, rs28676508, rs28632197, rs33985287), we reported both alleles and considered this site polymorphic. In order to infer the ancestral allele for these specific sites, we aligned the aforementioned SNPs with the orangutan (*Pongo abelii*) genome (PPYG2version) and the gorilla (*Gorilla gorilla gorilla*) genome (gorGor4 version) through Ensembl[114]. We used the same database when we wanted to assess the state of a specific variant in the rest of primates in the cases of convergence between modern humans and bonobo.

We then classified the alleles in evolutionary stages based on their distribution (presence or absence) in the different species/populations studied (e.g. Homo-specific, modern human-specific, Altai Neanderthal-specific). We then reviewed exhaustively the clinical significance of each one of these SNPs in present-day human populations. The literature filtering was performed through the Viewer tool of the National Center for Biotechnology Information[120]. SNPs not known to be related to social cognition, social disorders or any other relevant information were discarded. Specifically, of the 3160 single nucleotide variants identified on the *OXTR*, only 55 are mentioned in the literature. Of those, we included 19 in our study (34,54%). Of the 1375 single nucleotide variants identified on *AVPR1A*, 10 are mentioned in the literature. Of those, we included 7 (70%). And of the 988 single nucleotide variants identified on *AVPR1B*, 14 are mentioned in the literature. Of those, we included 3 (21,42%). The reader can find a full list of the SNPs that have been identified in modern humans on the genes studied, as well as a list of the archaic-specific polymorphisms known to date in the Supplementary Material (Tables S3-4).

In addition, we performed a transcription factor binding site prediction test using Lasagna2.0[121], and functional effects tests of exon variants with SNAP2[122] to all the variant-changes we had identified between modern and archaic human sequences. The Lasagna2.0 test did not yield any results.

We also multialigned all the gene sequences (*OXTR*, *AVPR1A* and *AVPR1B*) using only the reference genome sequences of the species included in the study: Human (GRCh38.p12), Neanderthal and Denisovan[11–13], the chimpanzee genome (Pan_tro_3.0), the bonobo genome (PANPAN1.1, Max-Planck Institute for Evolutionary Anthropology version) and the rhesus macaque genome (Mmul_8.0.1) publicly provided by Ensembl[114] (Suppl. Material).

We also included in our analysis several *AVPR1A*-microsatellites that have been associated with social-related phenotypes in the literature. More specifically we added as a sequence-search track the modern human RS3-$(CT)_4TT(CT)_8(GT)_{24}$, RS1-$(GATA)_{14}$, GT_{25} and the intronic AVR-$(GT)_{14}(GA)_{13}(A)_8$ microsatellite-sequences on the jbrowser (https://bioinf.eva.mpg.de/jbrowse/) and on the Integrative Genomics Viewer and looked for any possible differences in the Neanderthal (Altai and Vindija) and the Denisovan sequences. We did not find any changes in these regions, hence we did not make any further mention to this in the Results.

Acknowledgements

We thank Thomas O'Rourke for comments on the manuscript. We also thank Erich D. Jarvis and Evan E. Eichler for guidance on primate variation data.

Funding statement

CB acknowledges the financial support from the Spanish Ministry of Economy and Competitiveness/FEDER funds (grant FFI2016-78034-C2-1-P), a Marie Curie International Reintegration Grant from the European Union (PIRG-GA-2009-256413), research funds from the Fundació Bosch i Gimpera, from the Generalitat de Catalunya (2017-SGR-341), and the MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas 4903 (Evolinguistics: JP17H06379). CTh acknowledges support from the Generalitat de Catalunya in the form of a doctoral (FI) fellowship. AA acknowledges financial support from the Spanish Ministry of Economy and Competitiveness and the European Social Fund (BES-2017-080366).

Author Contributions Statement

CTh conceptualized and designed the study. AA ran all the tests, did the literature mining, generated the figures, and tables; CTh and AA ran the multialignment and handled the primate SNV-data; CB coordinated the study; CTh, AA and CB wrote the paper.

Data availability statement

All data generated or analysed during this study are included in the published version of the article. Its Supplementary Information files can be accessed at this link: https://www.biorxiv.org/content/biorxiv/suppl/2018/11/04/460584.DC1/460584-1.pdf

Additional Information

Competing interests

There is NO Competing financial or non-financial interest.

References

1. Meyer-Lindenberg, A., Domes, G., Kirsch, P. & Heinrichs, M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat. Rev. Neurosci.* **12**, 524–538 (2011).
2. Hoyle, C. H. Neuropeptide families and their receptors: evolutionary perspectives. *Brain Res.* **848**, 1–25 (1999). URL https://doi.org/10.1016/s0006-8993(99)01975-7. DOI 10.1016/s0006-8993(99)01975-7.
3. Song, Z. & Albers, H. E. Cross-talk among oxytocin and arginine-vasopressin receptors: Relevance for basic and clinical studies of the brain and periphery. *Front. Neuroendocrinol.* (2017). URL http://www.sciencedirect.com/science/article/pii/S009130221730064X. DOI https://doi.org/10.1016/j.yfrne.2017.10.004.
4. Donaldson, Z. R. & Young, L. J. Oxytocin, vasopressin, and the neurogenetics of sociality. *Sci.* **322**, 900–904 (2008). URL https://doi.org/10.1126/science.1158668. DOI 10.1126/science.1158668.
5. Israel, S. *et al.* Molecular genetic studies of the arginine vasopressin 1a receptor (avpr1a) and the oxytocin receptor (oxtr) in human behaviour: from autism to altruism with some notes in between. *Elsevier* **170**, 435 – 449 (2008). URL http://www.sciencedirect.com/science/article/pii/S0079612308004342. DOI https://doi.org/10.1016/S0079-6123(08)00434-2.
6. Dunbar, R. I. M. *Human evolution* (Oxford University Press, 2016).
7. Tomasello, M. The ultra-social animal. *Eur. J. Soc. Psychol.* **44**, 187–194 (2014). URL https://doi.org/10.1002/ejsp.2015. DOI 10.1002/ejsp.2015.
8. Theofanopoulou, C., Boeckx, C. & Jarvis, E. D. A hypothesis on a role of oxytocin in the social mechanisms of speech and vocal learning. *Proc. Royal Soc. B: Biol. Sci.* **284**, 20170988 (2017). URL https://doi.org/10.1098/rspb.2017.0988. DOI 10.1098/rspb.2017.0988.
9. Staes, N., Bradley, B. J., Hopkins, W. D. & Sherwood, C. C. Genetic signatures of socio-communicative abilities in primates. *Curr. Opin. Behav. Sci.* **21**, 33–38 (2018). URL https://doi.org/10.1016/j.cobeha.2017.11.013. DOI 10.1016/j.cobeha.2017.11.013.
10. Raghanti, M. A. *et al.* A neurochemical hypothesis for the origin of hominids. *Proc. Natl. Acad. Sci.* 201719666 (2018). URL https://doi.org/10.1073/pnas.1719666115. DOI 10.1073/pnas.1719666115.
11. Prüfer, K. *et al.* The complete genome sequence of a neanderthal from the altai mountains. *Nat.* **505**, 43–49 (2014). URL https://doi.org/10.1038/nature12886. DOI 10.1038/nature12886.
12. Prüfer, K. *et al.* A high-coverage neandertal genome from vindija cave in croatia. *Sci.* (2017). URL http://science.sciencemag.org/content/early/2017/10/04/science.aao1887. DOI 10.1126/science.aao1887. http://science.sciencemag.org/content/early/2017/10/04/science.aao1887.full.pdf.
13. Meyer, M. *et al.* A high-coverage genome sequence from an archaic denisovan individual. *Sci.* **338**, 222–226 (2012). URL http://science.sciencemag.org/content/338/6104/222. DOI 10.1126/science.1224344. http://science.sciencemag.org/content/338/6104/222.full.pdf.
14. Hare, B. Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annu. Rev. Psychol.* **68**, 155–186 (2017). URL https://doi.org/10.1146/annurev-psych-010416-044201. DOI 10.1146/annurev-psych-010416-044201. PMID: 27732802, https://doi.org/10.1146/annurev-psych-010416-044201.
15. Theofanopoulou, C. *et al.* Self-domestication in homo sapiens: Insights from comparative genomics. *PLOS ONE* **12**, e0185306 (2017). URL https://doi.org/10.1371/journal.pone.0185306. DOI 10.1371/journal.pone.0185306.

16. Pearce, E., Wlodarski, R., Machin, A. & Dunbar, R. I. M. Variation in the *beta*-endorphin, oxytocin, and dopamine receptor genes is associated with different dimensions of human sociality. *Proc. Natl. Acad. Sci.* **114**, 5300–5305 (2017). URL https://doi.org/10.1073/pnas.1700712114. DOI 10.1073/pnas.1700712114.

17. Kuhlwilm, M. & Boeckx, C. Genetic differences between humans and other hominins contribute to the "human condition". *bioRxiv* (2018). URL https://www.biorxiv.org/content/early/2018/04/11/298950. DOI 10.1101/298950. https://www.biorxiv.org/content/early/2018/04/11/298950.full.pdf.

18. Wade, M., Hoffmann, T. J. & Jenkins, J. M. Gene-environment interaction between the oxytocin receptor (OXTR) gene and parenting behaviour on children's theory of mind. *Soc. Cogn. Affect. Neurosci.* **10**, 1749–1757 (2014). DOI 10.1093/scan/nsv064.

19. Kawamura, Y. *et al.* The association between oxytocin receptor gene (OXTR) polymorphisms and affective temperaments, as measured by TEMPS-A. *J. Affect. Disord.* **127**, 31–37 (2010). URL http://dx.doi.org/10.1016/j.jad.2010.04.014. DOI 10.1016/j.jad.2010.04.014.

20. Schaschl, H. *et al.* Signatures of positive selection in the cis-regulatory sequences of the human oxytocin receptor (OXTR) and arginine vasopressin receptor 1a (AVPR1A) genes. *BMC Evol. Biol.* **15**, 85 (2015). URL http://www.biomedcentral.com/1471-2148/15/85. DOI 10.1186/s12862-015-0372-7.

21. Myers, A. J. *et al.* Variation in the oxytocin receptor gene is associated with increased risk for anxiety, stress and depression in individuals with a history of exposure to early life stress. *J Psychiatr Res* **59**, 93–100 (2014). DOI 10.1016/j.jpsychires.2014.08.021.VARIATION.

22. Wu, N., Li, Z. & Su, Y. The association between oxytocin receptor gene polymorphism (OXTR) and trait empathy. *J. Affect. Disord.* **138**, 468–472 (2012). URL http://dx.doi.org/10.1016/j.jad.2012.01.009. DOI 10.1016/j.jad.2012.01.009.

23. Wu, N. & Su, Y. Variations in the oxtr gene and prosocial behavior: Moderating effects of situational factors. *Integr. zoology* (2018).

24. Kang, J. I., Kim, H. W., Kim, C. H., Hwang, E. H. & Kim, S. J. Oxytocin receptor gene polymorphisms exert a modulating effect on the onset age in patients with obsessive-compulsive disorder. *Psychoneuroendocrinology* **86**, 45–52 (2017).

25. Park, J. *et al.* Evidence that genetic variation in the oxytocin receptor (OXTR) gene influences social cognition in ADHD. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **34**, 697–702 (2010). URL http://www.ncbi.nlm.nih.gov/pubmed/20347913http://dx.doi.org/10.1016/j.pnpbp.2010.03.029. DOI 10.1016/j.pnpbp.2010.03.029.

26. Baribeau, D. A. *et al.* Oxytocin receptor polymorphisms are differentially associated with social abilities across neurodevelopmental disorders. *Sci. reports* **7**, 11618 (2017).

27. Lerer, E. *et al.* Association between the oxytocin receptor (OXTR) gene and autism: Relationship to Vineland Adaptive Behavior Scales and cognition. *Mol. Psychiatry* **13**, 980–988 (2008). DOI 10.1038/sj.mp.4002087.

28. Schneiderman, I., Kanat-Maymon, Y., Ebstein, R. P. & Feldman, R. Cumulative risk on the oxytocin receptor gene (OXTR) underpins empathic communication difficulties at the first stages of romantic love. *Soc. Cogn. Affect. Neurosci.* **9**, 1524–1529 (2014). DOI 10.1093/scan/nst142.

29. Tansey, K. E. *et al.* Oxytocin receptor (OXTR) does not play a major role in the aetiology of autism: Genetic and molecular studies. *Neurosci. Lett.* **474**, 163–167 (2010). URL http://dx.doi.org/10.1016/j.neulet.2010.03.035. DOI 10.1016/j.neulet.2010.03.035.

30. Israel, S. *et al.* The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the Dictator Game and the social value orientations task. *PLoS One* **4** (2009). DOI 10.1371/journal.pone.0005535.

31. Chang, S.-C. *et al.* Are genetic variations in oxtr, avpr1a, and cd38 genes important to social integration? results from two large us cohorts. *Psychoneuroendocrinology* **39**, 257–268 (2014).

32. Harrison, A. J., Gamsiz, E. D., Berkowitz, I. C., Nagpal, S. & Jerskey, B. A. Genetic variation in the oxytocin receptor gene is associated with a social phenotype in autism spectrum disorders. *Am. J. Med. Genet. Part B: Neuropsychiatr. Genet.* **168**, 720–729 (2015).

33. Smearman, E. L. *et al.* Oxytocin receptor genetic and epigenetic variations: association with child abuse and adult psychiatric symptoms. *Child development* **87**, 122–134 (2016).

34. Sugar, C. A. & Green, M. F. Cognitive Performance in Individuals With Schizophrenia. *Schizophr Res* **159**, 353–357 (2015). DOI 10.1016/j.schres.2014.09.006.Associations.

35. Malik, A. I., Zai, C. C., Abu, Z., Nowrouzi, B. & Beitchman, J. H. The role of oxytocin and oxytocin receptor gene variants in childhood-onset aggression. *Genes, Brain Behav.* **11**, 545–551 (2012). DOI 10.1111/j.1601-183X.2012.00776.x.

36. Liu, X. *et al.* Association of the oxytocin receptor (OXTR) gene polymorphisms with autism spectrum disorder (ASD) in the Japanese population. *J. Hum. Genet.* **55**, 137–141 (2010). URL http://dx.doi.org/10.1038/jhg.2009.140. DOI 10.1038/jhg.2009.140.

37. Montag, C. *et al.* Association between Oxytocin Receptor Gene Polymorphisms and Self-Rated 'Empathic Concern' in Schizophrenia. *PLoS One* **7** (2012). DOI 10.1371/journal.pone.0051882.

38. Zhang, Y. *et al.* Genetic variants in oxytocin receptor gene (oxtr) and childhood physical abuse collaborate to modify the risk of aggression in chinese adolescents. *J. affective disorders* **229**, 105–110 (2018).

39. Feldman, R. *et al.* Sensitive parenting is associated with plasma oxytocin and polymorphisms in the OXTR and CD38 genes. *Biol. Psychiatry* **72**, 175–181 (2012). URL https://doi.org/10.1016/j.biopsych.2011.12.025. DOI 10.1016/j.biopsych.2011.12.025.

40. Ribeiro, L. d. O. P. *et al.* Evidence for association between oxtr gene and asd clinical phenotypes. *J. Mol. Neurosci.* 1–9 (2018).

41. Campbell, D. B. *et al.* Association of oxytocin receptor (OXTR) gene variants with multiple phenotype domains of autism spectrum disorder. *J. Neurodev. Disord.* **3**, 101–112 (2011). DOI 10.1007/s11689-010-9071-2.

42. Apicella, C. L. *et al.* No Association between Oxytocin Receptor (OXTR) Gene Polymorphisms and Experimentally Elicited Social Preferences. *PLoS One* **5**, 1–8 (2010). URL https://doi.org/10.1371/journal.pone.0011153. DOI 10.1371/journal.pone.0011153.

43. Dobewall, H. *et al.* Oxytocin receptor gene (oxtr) variant rs1042778 moderates the influence of family environment on changes in perceived social support over time. *J. affective disorders* **235**, 480–488 (2018).

44. Isgett, S. F., Algoe, S. B., Boulton, A. J., Way, B. M. & Fredrickson, B. L. Common variant in oxtr predicts growth in positive emotions from loving-kindness training. *Psychoneuroendocrinology* **73**, 244–251 (2016).

45. Wu, S. *et al.* Positive association of the oxytocin receptor gene (OXTR) with autism in the Chinese Han population. *Biol. Psychiatry* **58**, 74–77 (2005). DOI 10.1016/j.biopsych.2005.03.013.

46. Loparo, D. & Waldman, I. D. The oxytocin receptor gene (OXTR) is associated with autism spectrum disorder: a meta-analysis. *Mol. Psychiatry* **20**, 640–646 (2014). URL http://dx.doi.org/10.1038/mp.2014.77. DOI 10.1038/mp.2014.77.

47. Jacob, S. *et al.* Association of the oxytocin receptor gene (OXTR) in caucasian children and adolescents with autism. *Neurosci. Lett.* **417**, 6–9 (2007). URL https://doi.org/10.1016/j.neulet.2007.02.001. DOI 10.1016/j.neulet.2007.02.001.

48. Smearman, E. L. *et al.* Oxytocin receptor genetic and epigenetic variation: association with child abuse and adult psychiatric symptoms. *Child Dev.* **87**, 122–134 (2016). DOI 1doi:10.1111/cdev.12493. 15334406.

49. Flasbeck, V., Moser, D., Kumsta, R. & Brüne, M. The oxtr single-nucleotide polymorphism rs53576 moderates the impact of childhood maltreatment on empathy for social pain in female participants: evidence for differential susceptibility. *Front. psychiatry* **9** (2018).

50. Tombeau Cost, K. *et al.* Thinking and doing: the effects of dopamine and oxytocin genes and executive function on mothering behaviours. *Genes, Brain Behav.* **16**, 285–295 (2017).

51. Chen, F. S. & Johnson, S. C. Oxytocin receptor (OXTR) polymorphisms and attachment in human infants. *Front. Psychol.* **2**, 1–6 (2011). DOI 10.3389/fpsyg.2011.00200.

52. Yang, S. *et al.* Serum oxytocin levels and an oxytocin receptor gene polymorphism (rs2254298) indicate social deficits in children and adolescents with autism spectrum disorders. *Front. neuroscience* **11**, 221 (2017).

53. Cattaneo, Z. *et al.* Congenital prosopagnosia is associated with a genetic variation in the oxytocin receptor (oxtr) gene: An exploratory study. *Neurosci.* **339**, 162–173 (2016).

54. Shao, D. *et al.* Effect of the interaction between oxytocin receptor gene polymorphism (rs53576) and stressful life events on aggression in chinese han adolescents. *Psychoneuroendocrinology* (2018).

55. Kim, H. W., Kang, J. I., An, S. K. & Kim, S. J. Oxytocin receptor gene variants are associated with emotion recognition and resilience, but not with false-belief reasoning performance in healthy young korean volunteers. *CNS neuroscience & therapeutics* (2018).

56. Schiele, M. A. *et al.* Oxytocin receptor gene variation, behavioural inhibition, and adult separation anxiety: Role in complicated grief. *The World J. Biol. Psychiatry* 1–9 (2018).

57. Andreou, D., Comasco, E., Åslund, C., Nilsson, K. W. & Hodgins, S. Maltreatment, the oxytocin receptor gene, and conduct problems among male and female teenagers. *Front. human neuroscience* **12**, 112 (2018).

58. Inoue, H. *et al.* Association between the oxytocin receptor gene and amygdalar volume in healthy adults. *Biol. Psychiatry* **68**, 1066–1072 (2010). URL http://dx.doi.org/10.1016/j.biopsych.2010.07.019. DOI 10.1016/j.biopsych.2010.07.019.

59. Marusak, H. A. *et al.* Amygdala responses to salient social cues vary with oxytocin receptor genotype in youth. *Neuropsychol.* **79**, 1–9 (2015).

60. Tost, H. *et al.* Neurogenetic effects of OXTR rs2254298 in the extended limbic system of healthy caucasian adults. *Biol. Psychiatry* **70**, 37–39 (2011). DOI 10.1016/j.biopsych.2011.06.034.

61. Nishina, K. *et al.* Association of the oxytocin receptor gene with attitudinal trust: role of amygdala volume. *Soc. cognitive affective neuroscience* **1**, 7 (2018).

62. Holmqvist Jämsen *et al.* Associations Between Vocal Symptoms and Genetic Variants in the Oxytocin Receptor and Arginine Vasopressin 1A Receptor Gene. *J. Speech Lang. Hear. Res.* **60**, 1843 (2017).

63. Francis, S. M. *et al.* ASD and genetic associations with receptors for oxytocin and vasopressin-AVPR1A, AVPR1B, and OXTR. *Front. Neurosci.* **10**, 1–10 (2016). DOI 10.3389/fnins.2016.00516.

64. Napoli, A. D., Warrier, V., Baron-Cohen, S. & Chakrabarti, B. Genetic variation in the oxytocin receptor (OXTR) gene is associated with asperger syndrome. *Mol. Autism* **5**, 48 (2014). URL https://doi.org/10.1186/2040-2392-5-48. DOI 10.1186/2040-2392-5-48.

65. Yrigollen, C. M. *et al.* Genes controlling affiliative behavior as candidate genes for autism. *Biol. Psychiatry* **63**, 911–916 (2008). URL https://doi.org/10.1016/j.biopsych.2007.11.015. DOI 10.1016/j.biopsych.2007.11.015.

66. Davis, M. C. *et al.* Associations between oxytocin receptor genotypes and social cognitive performance in individuals with schizophrenia. *Schizophr. Res.* **159**, 353–357 (2014). URL https://doi.org/10.1016/j.schres.2014.09.006. DOI 10.1016/j.schres.2014.09.006.

67. Ayaz, A. B. *et al.* Oxytocin system social function impacts in children with attention-deficit/hyperactivity disorder. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **168**, 609–616 (2015). DOI 10.1002/ajmg.b.32343.

68. Damiano, C. R. *et al.* Association between the oxytocin receptor (OXTR) gene and mesolimbic responses to rewards. *Mol. Autism* **5**, 7 (2014). URL http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-5-7. DOI 10.1186/2040-2392-5-7.

69. Bakermans-Kranenburg, M. J. & Van Ijzendoorn, M. H. A sociability gene' Meta-Analysis of oxytocin receptor genotype effects in humans. *Psychiatr. Genet.* **24**, 45–51 (2014). DOI 10.1097/YPG.0b013e3283643684.

70. Bernhard, R. M. *et al.* Variation in the oxytocin receptor gene (OXTR) is associated with differences in moral judgment. *Soc. Cogn. Affect. Neurosci.* nsw103 (2016). URL https://academic.oup.com/scan/article-lookup/doi/10.1093/scan/nsw103. DOI 10.1093/scan/nsw103.

71. Tabak, B. A., McCullough, M. E., Carver, C. S., Pedersen, E. J. & Cuccaro, M. L. Variation in oxytocin receptor gene (OXTR) polymorphisms is associated with emotional and behavioral reactions to betrayal. *Soc. Cogn. Affect. Neurosci.* **9**, 810–816 (2013). DOI 10.1093/scan/nst042.

72. Chang, S.-C. *et al.* Are genetic variations in OXTR, AVPR1a, and CD38 genes important to social integration? results from two large u.s. cohorts. *Psychoneuroendocrinology* **39**, 257–268 (2014). URL https://doi.org/10.1016/j.psyneuen.2013.09.024. DOI 10.1016/j.psyneuen.2013.09.024.

73. Kim, Y.-R., Kim, J.-H., Kim, C.-H., Shin, J. G. & Treasure, J. Association between the Oxytocin Receptor Gene Polymorphism (rs53576) and Bulimia Nervosa. *Eur. Eat. Disord. Rev.* **23**, 171–178 (2015). URL http://doi.wiley.com/10.1002/erv.2354. DOI 10.1002/erv.2354.

74. Chen, F. S. *et al.* Common oxytocin receptor gene (OXTR) polymorphism and social support interact to reduce stress in humans. *Proc. Natl. Acad. Sci.* **108**, 19937–19942 (2011). URL https://doi.org/10.1073/pnas.1113079108. DOI 10.1073/pnas.1113079108.

75. Costa, B. *et al.* Oxytocin receptor polymorphisms and adult attachment style in patients with depression. *Psychoneuroendocrinology* **34**, 1506–1514 (2009). URL https://doi.org/10.1016/j.psyneuen.2009.05.006. DOI 10.1016/j.psyneuen.2009.05.006.

76. Feng, C. *et al.* A common oxytocin receptor gene (OXTR) polymorphism modulates intranasal oxytocin effects on the neural response to social cooperation in humans. *Genes, Brain Behav.* **14**, 516–525 (2015). DOI 10.1111/gbb.12234.

77. Laursen, H. R. *et al.* Variation in the oxytocin receptor gene is associated with behavioral and neural correlates of empathic accuracy. *Front. Behav. Neurosci.* **8** (2014). URL https://doi.org/10.3389/fnbeh.2014.00423. DOI 10.3389/fnbeh.2014.00423.

78. Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P. & Keltner, D. Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proc. Natl. Acad. Sci.* **106**, 21437–21441 (2009). URL https://doi.org/10.1073/pnas.0909579106. DOI 10.1073/pnas.0909579106.

79. Saphire-Bernstein, S., Way, B. M., Kim, H. S., Sherman, D. K. & Taylor, S. E. Oxytocin receptor gene (OXTR) is related to psychological resources. *Proc. Natl. Acad. Sci.* **108**, 15118–15122 (2011). URL https://doi.org/10.1073/pnas.1113137108. DOI 10.1073/pnas.1113137108.

80. McQuaid, R. J., McInnis, O. A., Stead, J. D., Matheson, K. & Anisman, H. A paradoxical association of an oxytocin receptor gene polymorphism: early-life adversity and vulnerability to depression. *Front. Neurosci.* **7** (2013). URL https://doi.org/10.3389/fnins.2013.00128. DOI 10.3389/fnins.2013.00128.

81. Chen, F. S. *et al.* Genetic modulation of oxytocin sensitivity: a pharmacogenetic approach. *Transl. Psychiatry* **5**, e664–e664 (2015). URL https://doi.org/10.1038/tp.2015.163. DOI 10.1038/tp.2015.163.

82. Lucht, M. J. *et al.* Associations between the oxytocin receptor gene (OXTR) and affect, loneliness and intelligence in normal subjects. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **33**, 860–866 (2009). URL http://dx.doi.org/10.1016/j.pnpbp.2009.04.004. DOI 10.1016/j.pnpbp.2009.04.004.

83. Lucht, M. J. *et al.* Associations between the oxytocin receptor gene (OXTR) and "mind-reading" in humans - An exploratory study. *Nord. J. Psychiatry* **67**, 15–21 (2013). URL http://www.tandfonline.com/doi/full/10.3109/08039488.2012.700731. DOI 10.3109/08039488.2012.700731.

84. Kalyoncu, T., Özbaran, B., Köse, S. & Onay, H. Variation in the oxytocin receptor gene is associated with social cognition and adhd. *J. attention disorders* 1087054717706757 (2017).

85. Tansey, K. E. *et al.* Functionality of promoter microsatellites of arginine vasopressin receptor 1a (AVPR1a): implications for autism. *Mol. Autism* **2**, 3 (2011). URL https://doi.org/10.1186/2040-2392-2-3. DOI 10.1186/2040-2392-2-3.

86. Kantojärvi, K. *et al.* Association and promoter analysis ofAVPR1ain finnish autism families. *Autism Res.* **8**, 634–639 (2015). URL https://doi.org/10.1002/aur.1473. DOI 10.1002/aur.1473.

87. Levran, O. *et al.* Stress-related genes and heroin addiction: A role for a functional FKBP5 haplotype. *Psychoneuroendocrinology* **45**, 67–76 (2014). URL https://doi.org/10.1016/j.psyneuen.2014.03.017. DOI 10.1016/j.psyneuen.2014.03.017.

88. Maher, B. S. *et al.* The AVPR1a gene and substance use disorders: Association, replication, and functional evidence. *Biol. Psychiatry* **70**, 519–527 (2011). URL https://doi.org/10.1016/j.biopsych.2011.02.023. DOI 10.1016/j.biopsych.2011.02.023.

89. Reuter, M., Cooper, A. J., Smillie, L. D., Markett, S. & Montag, C. A new measure for the revised reinforcement sensitivity theory: psychometric criteria and genetic validation. *Front. Syst. Neurosci.* **9** (2015). URL https://doi.org/10.3389/fnsys.2015.00038. DOI 10.3389/fnsys.2015.00038.

90. Saccone, S. F. *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* **16**, 36–49 (2006). URL https://doi.org/10.1093/hmg/ddl438. DOI 10.1093/hmg/ddl438.

91. Levran, O. *et al.* Heroin addiction in african americans: a hypothesis-driven association study. *Genes, Brain Behav.* **8**, 531–540 (2009). URL https://doi.org/10.1111/j.1601-183x.2009.00501.x. DOI 10.1111/j.1601-183x.2009.00501.x.

92. Yang, S. Y. *et al.* Association study between single nucleotide polymorphisms in promoter region of AVPR1a and korean autism spectrum disorders. *Neurosci. Lett.* **479**, 197–200 (2010). URL https://doi.org/10.1016/j.neulet.2010.05.050. DOI 10.1016/j.neulet.2010.05.050.

93. Yang, S. Y. *et al.* Replicative genetic association study between functional polymorphisms in AVPR1a and social behavior scales of autism spectrum disorder in the korean population. *Mol. Autism* **8** (2017). URL https://doi.org/10.1186/s13229-017-0161-9. DOI 10.1186/s13229-017-0161-9.

94. Zai, C. C. *et al.* Possible genetic association between vasopressin receptor 1b and child aggression. *Psychiatry Res.* **200**, 784–788 (2012). URL https://doi.org/10.1016/j.psychres.2012.07.031. DOI 10.1016/j.psychres.2012.07.031.

95. Keck, M. E. *et al.* Combined effects of exonic polymorphisms in CRHR1 and AVPR1b genes in a case/control study for panic disorder. *Am. J. Med. Genet. Part B: Neuropsychiatr. Genet.* **147B**, 1196–1204 (2008). URL https://doi.org/10.1002/ajmg.b.30750. DOI 10.1002/ajmg.b.30750.

96. Dempster, E. L. Evidence of an association between the vasopressin v1b receptor gene (AVPR1b) and childhood-onset mood disorders. *Arch. Gen. Psychiatry* **64**, 1189 (2007). URL https://doi.org/10.1001/archpsyc.64.10.1189. DOI 10.1001/archpsyc.64.10.1189.

97. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nat.* **499**, 471 (2013).

98. Staes, N. *et al.* Chimpanzee sociability is associated with vasopressin (avpr1a) but not oxytocin receptor gene (oxtr) variation. *Horm. behavior* **75**, 84–90 (2015).

99. Staes, N. *et al.* Bonobo personality traits are heritable and associated with vasopressin receptor gene 1a variation. *Sci. reports* **6**, 38193 (2016).

100. Curley, J. P. & Keverne, E. B. Genes, brains and mammalian social bonds. *Trends Ecol. & Evol.* **20**, 561–567 (2005). URL https://doi.org/10.1016/j.tree.2005.05.018. DOI 10.1016/j.tree.2005.05.018.

101. O'Rourke, T. & Boeckx, C. Converging roles of glutamate receptors in domestication and prosociality. *bioRxiv* (2018). URL https://www.biorxiv.org/content/early/2018/10/11/439869. DOI 10.1101/439869.

102. Schaefer, K., Fink, B., Mitteroecker, P., Neave, N. & Bookstein, F. L. Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. *Coll. antropologicum* **29**, 415–419 (2005).

103. Nelson, E., Rolian, C., Cashmore, L. & Shultz, S. Digit ratios predict polygyny in early apes, ardipithecus, neanderthals and early modern humans but not in australopithecus. *Proc. Royal Soc. Lond. B: Biol. Sci.* **278**, 1556–1563 (2011).

104. Weisman, O. *et al.* The association between 2d: 4d ratio and cognitive empathy is contingent on a common polymorphism in the oxytocin receptor gene (oxtr rs53576). *Psychoneuroendocrinology* **58**, 23–32 (2015).

105. Fam, B. S. *et al.* Oxytocin and arginine vasopressin systems in the domestication process. *Genet.* **41**, 235 – 242 (2018). URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572018000200235&nrm=iso.

106. Herbeck, Y. E. & Gulevich, R. G. Neuropeptides as facilitators of domestication. *Cell Tissue Res.* 1–13 (2018).

107. Tan, J., Ariely, D. & Hare, B. Bonobos respond prosocially toward members of other groups. *Sci. Reports* **7** (2017). URL https://doi.org/10.1038/s41598-017-15320-w. DOI 10.1038/s41598-017-15320-w.

108. Hare, B. & Wrangham, R. W. Equal, similar, but different: Convergent bonobos and conserved chimpanzees. In Muller, M. M., Wrangham, R. & Pilbeam, D. R. (eds.) *Chimpanzees and Human Evolution*, chap. 3, 142–173 (Harvard University Press, Harvard, 2017).

109. Chamary, J., Parmley, J. L. & Hurst, L. D. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**, 98 (2006).

110. Rogers, A. R., Bohlender, R. J. & Huff, C. D. Early history of neanderthals and denisovans. *Proc. Natl. Acad. Sci.* **114**, 9859–9863 (2017). URL https://doi.org/10.1073/pnas.1706426114. DOI 10.1073/pnas.1706426114.

111. Belsky, J. Variation in susceptibility to environmental influence: An evolutionary argument. *Psychol. inquiry* **8**, 182–186 (1997).

112. Hublin, J.-J., Neubauer, S. & Gunz, P. Brain ontogeny and life history in pleistocene hominins. *Phil. Trans. R. Soc. B* **370**, 20140062 (2015).

113. Auton, A. *et al.* A global reference for human genetic variation. *Nat.* **526**, 68–74 (2015). URL https://doi.org/10.1038/nature15393. DOI 10.1038/nature15393.

114. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2015). URL https://doi.org/10.1093/nar/gkv1157. DOI 10.1093/nar/gkv1157.

115. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinforma.* **30**, 3276–3278 (2014). URL https://doi.org/10.1093/bioinformatics/btu531. DOI 10.1093/bioinformatics/btu531.

116. Wright, E. Decipher (2017). DOI 10.18129/b9.bioc.decipher.

117. Edgar, R. C. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).

118. Brown, N. P., Leroy, C. & Sander, C. Mview: a web-compatible database search or multiple alignment viewer. *Bioinforma. (Oxford, England)* **14**, 380–381 (1998).

119. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. biotechnology* **29**, 24 (2011).

120. NCBI Resource Coordinators. Database resources of the ncbi. *Nucleic Acids Res.* **45**, D12–D17 (2016). URL https://doi.org/10.1093/nar/gkw1071. DOI 10.1093/nar/gkw1071.

121. Lee, C. & Huang, C.-H. LASAGNA-search: an integrated web tool for transcription factor binding site search and visualization. *BioTechniques* **54** (2013). URL https://doi.org/10.2144/000113999. DOI 10.2144/000113999.

122. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC Genomics* **16**, S1 (2015). URL https://doi.org/10.1186/1471-2164-16-s8-s1. DOI 10.1186/1471-2164-16-s8-s1.

Appendix

Chapter 2

Selective vocal learning in a social reward context

Theofanopoulou C. [1,2], Lipkind D. [5], Tchernichovski O. [5], Boeckx C. [1,3] & Jarvis E.D. [2,4]

[1]Section of General Linguistics, University of Barcelona, Barcelona, Spain

[2]Department of Neurogenetics of Language, Rockefeller University, New York, USA

[3]ICREA

[4]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

[5]City University of New York - Hunter College, New York, USA

Abstract

Social reward has been traditionally thought to enhance learning, with most experiments testing whether it makes learning faster or better. It remains unclear how social reward affects spoken language. We hypothesized that social reward affects a specialized component of spoken language, vocal learning. We tested this hypothesis using the zebra finch (*Taeniopygia guttata*), a vocal learning songbird commonly used as a model for human spoken-language development. To do so, we developed a rapid vocal learning behavioral paradigm that attempted to dissociate social reward from vocal learning. Juvenile male zebra finches were first operantly taught to imitate a two-syllable song for 20 days. Then for the next for 30 days they were exposed to two different contexts, switched every other day: an isolation context and a social reward context with an animal model of a bird they treated as their father, and a non-singing but live female bird. In both contexts, they were exposed to operantly elicited playbacks of one of two very similar songs, comprised of two syllables, the same syllables of the song they had learnt, only differing by two semitones in the pitch of the second syllable. Five out of the six birds tested imitated the pitch of the song they heard in the social reward context, suggesting that fine-grained aspects of vocal learning, like pitch, can be gated by social reward. Given the convergent behaviors and neural pathways for learned vocal communication in zebra finches and humans, our results imply that social reward could help gate learning of features of speech.

Introduction

Early language acquisition is strongly shaped by social interactions, as seen in qualitative observations of every day experience and by quantitative scientific studies. Kuhl 2007 showed that when human infants were exposed to only audiovisual or audio recordings of a language, they would not learn it successfully, whereas they would learn it if it was followed by live social feedback. Although these studies point to a crucial role of social interactions in language acquisition, they were conducted for a parallel second language which would be acquired through these experiments along with the first language that infants were learning in the natural way; so it remains to be elucidated whether the same would hold for their first language. Other studies have highlighted the importance of social feedback and contingency in first language acquisition by showing that a child's vocalization tends to be speech-related only if the previous speech-related vocalization received an immediate adult feedback (Warlaumont et al. 2014).

We hypothesize that such feedback is a social reward that vigorously affects fine sensory-motor aspects of a specialized component of speech/language learning, called vocal learning. Vocal learning is the ability to imitate sounds, found to date in only a few independently evolved species of mammals (humans, bats, cetaceans, sea lions and elephants) and birds (songbirds, parrots and hummingbirds) (Janik and Slater 1997; Jarvis 2004). Of the non-human vocal learners, songbirds have been studied the most. Their vocal-learning ability displays parallels to human speech learning, having undergone convergent evolution, at the level of behavior, neural connectivity and gene expression specializations in song and speech brain regions

(Doupe and Kuhl 1999; Jarvis 2004; Pfenning et al. 2014). Thus, by dissecting the social mechanisms of vocal learning in songbirds, we may illuminate how social interactions shape vocal learning in humans.

The thoroughly studied and highly gregarious zebra finch figures as a promising candidate for modelling the impact that social factors have on human vocal learning. Specifically, in laboratory tests, juvenile zebra finches have been shown to learn best from a live tutor (Eales 1989), while learning from purely tape recorded songs is less effective (Beecher 2017).

In this study, we sought to tease apart, if possible, vocal learning from social reward, and then test their interactions. We developed a learning paradigm with juvenile zebra finches, where we taught them to learn a source two-syllable song in the first days of their lives in one context (day 40-60), and then in days 60-90 we switched them to two different contexts, one with a 'social reward' and the other in 'isolation'. In these two different contexts, they were exposed to two very similar songs (played from speakers), only differing by two semitones in the pitch of the second syllable of the original song they learned. We found that birds modified their songs to the playbacks heard, with most of them crystallizing on the syllable acoustic features heard in the more highly social context. Our findings are the first we are aware of experimentally showing, in the same individual, that social reward can motivate vocal imitation of sounds heard.

Methods

Animals

Animal care and experimental procedures were approved by the Institutional Animal Care and Use Committee of Hunter College. We used male and female zebra finches from the breeding colony at Hunter College.

Vocal learning and social reward paradigm

We developed a vocal learning and social reward paradigm that has two Phases (Figure 1). For Phase I, we followed a protocol similar to Lipkind et al. 2017. Male zebra finches were bred and reared in cages in sound isolation boxes with their mother (who does not produce learned song), but in the absence of their father and other adult males (that do produce learned song), between days 7–30 post hatch. Afterwards, birds were removed from their home cage and were kept singly (without their mother) in sound attenuation chambers, and continuously recorded. Throughout the experiment birds were on a 12-hour light and 12-hour dark schedule. From days 30-40, birds were passively exposed to 20 playbacks per day (10 in the morning and 10 in the afternoon) of a two-syllable song (AB), occurring at random with a probability of 0.005 per second ('passive' training) (Figure 2a, 2c). On day 40, birds were trained to press a key to hear song playbacks, with a daily quota of 20 ('active' training). During 'active' training (days 40-60) we placed in each bird's cage an animal model of an adult male bird that we believe they treated as substitute for their father (Figure 2b). Previous experiments have shown that the presence of an adult male model animal, along with the active pecking that induces the playback, leads to faster learning and a more accurate copy of the source song (Tchernichovski et al. 2001; Derégnaucourt et al. 2013). Only the birds who had successfully learnt the source song (AB) until day 60 were selected to be passed onto Phase II. Learning of the source was assessed by quantifying the percent similarity (Sound Analysis Pro; Tchernichovski et al. 2000; Tchernichovski et al. 2004) between the bird's song motifs and the source model motif in 10 randomly chosen song bouts per day. We considered the source song as learnt when the acoustic similarity to the model was at least 70%. We paid particular attention to the pitch of the second syllable (B; ~1160 Hz frequency), since this would be the feature that we would tweak in Phase II.

For Phase II, the juvenile males were exposed to two different contexts, switched every other day: an isolation context and a social reward context. The birds spent 24 hours in each context and were switched to the other context in the morning after the lights in their cages went on. As in Phase I, the daily quota of playbacks was 20 per day in either contexts to ensure equal exposure. In the isolation context, birds were put alone in sound isolation chambers, while they could still peck on the keys and induce playbacks from the speakers (Figure 3). In the social reward context, birds were housed with an animal model of a bird they treated as their father (the same we had used from day 40-60), a non-singing but live female bird and a mirror which gives them the illusion there are more birds around. In this context, the birds still had to peck on a key to induce playbacks from the speakers (Figure 3). In these two different contexts, they were exposed to two very similar songs (played from speakers), comprised of two syllables, the same syllables of the song they had learnt (AB), but differing by two semitones in the pitch of the second syllable B. This was either shifting the pitch by two semitones up (ABplus; AB+) or two semitones down (ABminus; AB-). B+'s fundamental frequency was 1300 Hz, while B-'s frequency was 1050 Hz (source B: 1160 Hz). AB- and AB+ were assigned to either the isolation or the social reward context randomly for each bird. We did not assign the same song (e.g. AB+) to the same context (e.g. social reward) for all birds, to make sure we avoid measuring a shift bias as opposed to a social bias. Birds were never put back into the main colony between training sessions or Phases.

Behavioral analyses

Recordings and trainings were done using Sound Analysis Pro (Tchernichovski et al. 2000; Tchernichovski et al. 2004), and continued until birds reached day 90, when they crystalize onto the song they will be singing for the rest of their lives (Zann 1996). Source and target song models were synthetically composed of natural syllables. Pitch mismatches between source and target syllables were generated with GOLDWAVE v. 5.68 (www.goldwave.com). Song feature calculation and cluster analysis were performed using Sound Analysis Pro, on a randomly selected 10% of the sound files in each developmental day. Cluster information was used to track changes in the spectral structure (specifically, median pitch) of the B syllable.

Results

During training, 6 juvenile male birds of 13 tested successfully learned the AB song in Phase 1. It took on average 25± days for them to reach the 70% syllable identity to the tutor song criterion. They did so on average day 57±.

In Phase 2, five out of the six birds that we tested switched the pitch of the B syllable to the pitch they heard while in the social reward context. In two of these cases, we had assigned B+ to the social reward context, and in three cases B-, which is what the birds ended up imitating. For the first 8-10 days of Phase II, birds generally kept singing AB (i.e. B in the source pitch), while they sang some instances of both AB+ and AB-. They would on some days sing more AB+ and others more AB-, gradually going back and forth as if experimenting. Then around day 70-75 all birds had stabilized their song to the pitch of their choice until the end of their developmental recording period at ~ day 90 (Figure 4).

We further examined the only bird that did not switch his pitch to the social reward context. This bird ended up singing AB+ heard in the isolation context, but also introduced to his song an additional syllable (B'; Figure 5), that was shifted by 2 octaves down from the actual original B syllable matching the pitch of the B- version. That is, his additional syllable almost matched the pitch of the B syllable in the social reward context, while for his original song B syllable he opted for the pitch heard in the social isolation context.

Discussion

In our study we found that all juvenile birds changed the pitch of their songs when placed in a new context during juvenile song learning development, where five out of six of them imitated the pitch they were listening to while in the social reward context. The bird that changed in the opposite direction, also added another syllable at a similar pitch to the one he heard in the social reward context, indicating that the latter context still influenced additional imitation. These findings suggest that the birds preferred to imitate the song they heard in the social reward-context.

One interpretation of our findings is that songs heard in a condition with live and model animals is rewarding to the developing juvenile. Another interpretation is that juveniles perceived the playback from the speakers they were exposed to in the social reward context as coming from the animal model that they treated as their father since Phase I, as opposed to a speaker in the isolation context. In either case, our findings indicate that a social association with the songs heard even after initial vocal learning has occurred, can influence fine-tuned learning of changes in song.

These social vocalizations are probably associated with activation of genes in the brain that strengthen the newly formed synaptic interactions to hold onto these social associated auditory memories, which eventually shape the vocal learning pathway. As we hypothesized in Theofanopoulou, Boeckx, and Jarvis 2017, oxytocin provides a good candidate molecule potentially involved in the mechanism that links the social motivation-circuitry and the core vocal learning circuitry. An alternative is dopamine, or oxytocin and dopamine synergistically.

This study is a first step for an explanatory basis of why speech pitch-perception (auditory) (Wang et al. 2017) and performance (vocal) (Bonneh et al. 2011) present abnormalities in cases of social deficits, like autism. It provides a window into early therapeutic approaches that focus more on social feedback than on sensory-motor aspects of vocal learning.

We conclude that social reward affects at least the fine-grained aspects of vocal learning, like pitch learning. Our future studies will focus on different spectral and temporal song-features and assess whether these features too are shaped by social reward, as well as testing the specific features of the social reward context that motivates the juvenile to learn song from conspecifics.

Figures

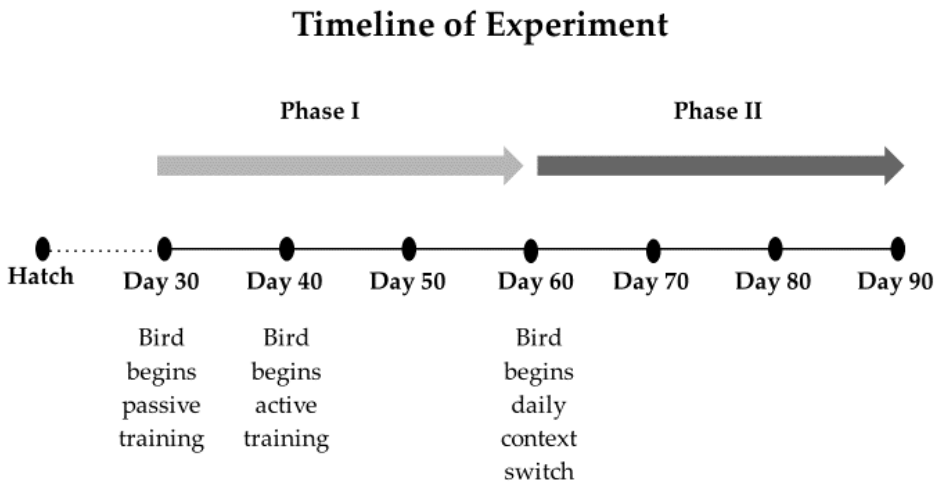


Figure 1: Timeline of the experiment

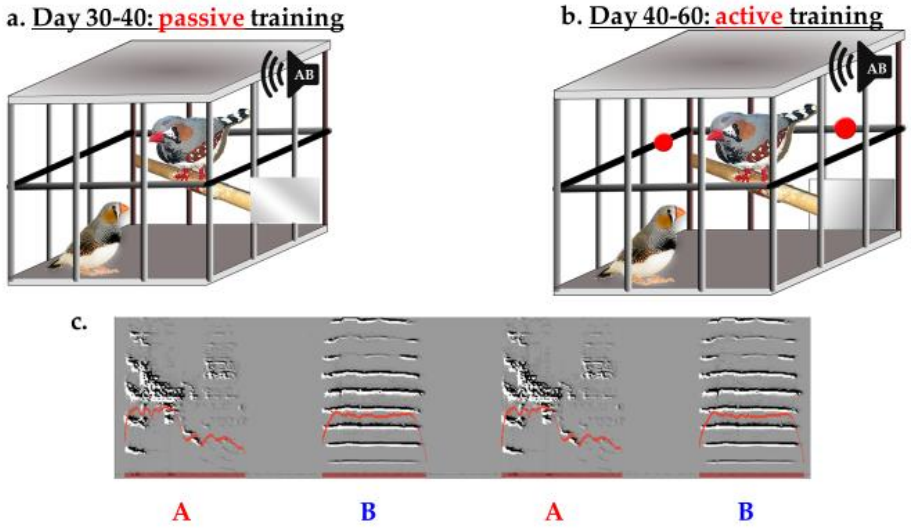


Figure 2: Phase 1 context. a. A representation of passive training of a juvenile live bird, in the presence of a model male animal, a mirror, and a source song-AB played from a speaker. **b.** A representation of active training of a juvenile bird, with a model animal, mirror, and keys that induce the source song to be played when pecked on. **c.** A spectrogram of the source song. Red bottom lines represent syllable boundaries; red upper lines show the amplitude of each syllable.

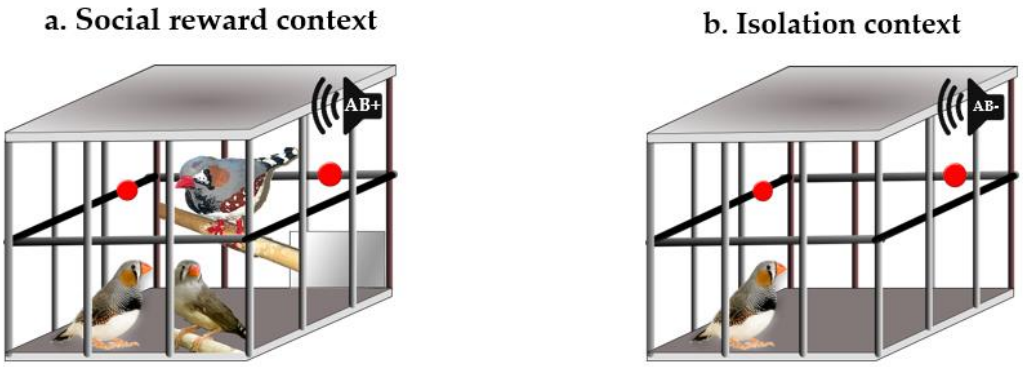


Figure 3: Phase II contexts. a. A representation of the social reward context with a juvenile-experimental male, a live non-singing female, a model animal, keys, speaker, and mirror. b. Isolation context with a juvenile-experimental male, keys, and speaker. We assigned AB+ to the social reward context and AB- to the isolation context for the shake of illustration.

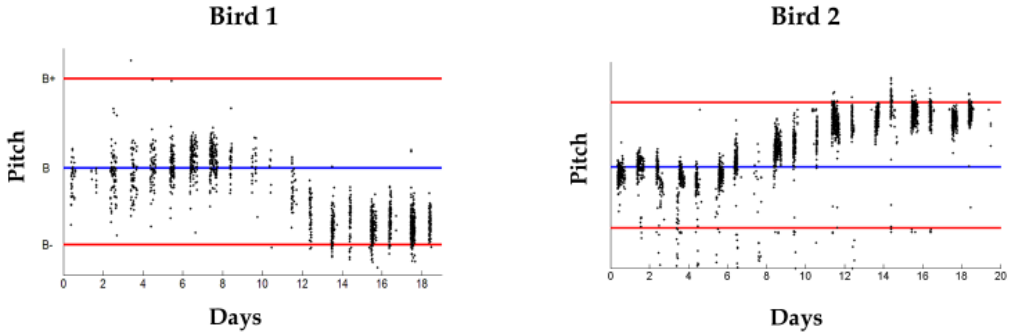


Figure 4: Pitch trajectories of two experimental birds that ended up picking up the pitch of syllable B that they listened to while in the social reward context: B- for Bird 1, B+ for Bird 2. In both cases the birds shifted their average pitch slightly up or down from days 0-8 in Phase II, but afterwards made a bigger shift in the social reward context direction. Day 0 equals the first day of Phase II, in other words Day 60 of the birds' life.

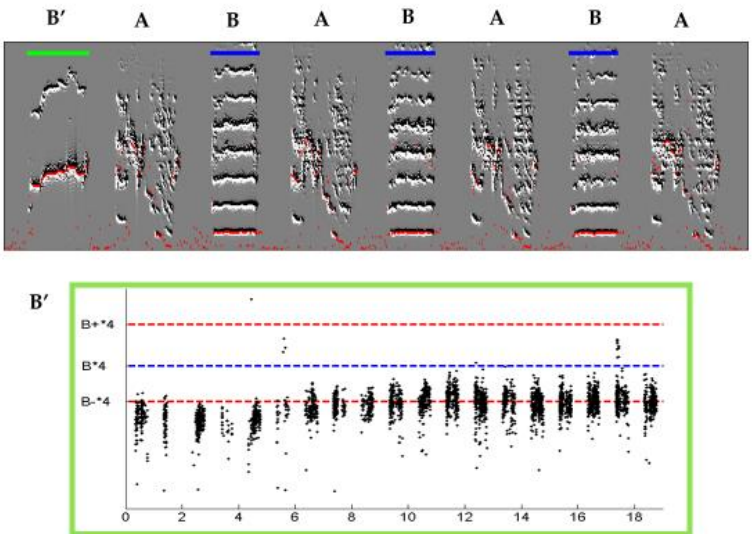


Figure 5: Example of a spectrogram of the song of the only bird that did not end up imitating the social reward pitch. B' (green-highlighted) is the additional syllable he introduced (top); a pitch-trajectory analysis of B' is shown at the bottom.

References

Beecher, Michael D. 2017. "Birdsong Learning as a Social Process." *Animal Behaviour* 124: 233–46. https://linkinghub.elsevier.com/retrieve/pii/S0003347216302020.

Bonneh, Yoram S. et al. 2011. "Abnormal Speech Spectrum and Increased Pitch Variability in Young Autistic Children." *Frontiers in Human Neuroscience* 4. http://journal.frontiersin.org/article/10.3389/fnhum.2010.00237/abstract.

Derégnaucourt, Sébastien et al. 2013. "Comparisons of Different Methods to Train a Young Zebra Finch (Taeniopygia Guttata) to Learn a Song." *Journal of Physiology-Paris* 107(3): 210–18. http://linkinghub.elsevier.com/retrieve/pii/S0928425712000435.

Doupe, Allison J., and Patricia K. Kuhl. 1999. "Birdsong and Human Speech: Common Themes and Mechanisms." *Annual Review of Neuroscience* 22(1): 567–631. http://www.annualreviews.org/doi/10.1146/annurev.neuro.22.1.567.

Eales, Lucy A. 1989. "The Influences of Visual and Vocal Interaction on Song Learning in Zebra Finches." *Animal Behaviour* 37: 507–8. http://linkinghub.elsevier.com/retrieve/pii/0003347289900973.

Janik, Vincent M., and Peter J.B. Slater. 1997. "Vocal Learning in Mammals." In , 59–99. http://linkinghub.elsevier.com/retrieve/pii/S0065345408603770.

Jarvis, Erich D. 2004. "Learned Birdsong and the Neurobiology of Human Language." *Annals of the New York Academy of Sciences* 1016(1): 749–77. http://doi.wiley.com/10.1196/annals.1298.038.

Kuhl, Patricia K. 2007. "Is Speech Learning 'Gated' by the Social Brain?" *Developmental Science* 10(1): 110–20. http://doi.wiley.com/10.1111/j.1467-7687.2007.00572.x.

Lipkind, Dina et al. 2017. "Songbirds Work around Computational Complexity by Learning Song Vocabulary Independently of Sequence." *Nature Communications* 8(1): 1247. http://www.nature.com/articles/s41467-017-01436-0.

Pfenning, A. R. et al. 2014. "Convergent Transcriptional Specializations in the Brains of Humans and Song-Learning Birds." *Science* 346(6215): 1256846–1256846.

http://www.sciencemag.org/cgi/doi/10.1126/science.1256846.

Tchernichovski, O. Mitra, P.P., Lints, T., Nottebohm, F. 2001. “Dynamics of the Vocal Imitation Process: How a Zebra Finch Learns Its Song.” *Science* 291(5513): 2564–69. http://www.sciencemag.org/cgi/doi/10.1126/science.1058522.

Tchernichovski, O. et al. 2004. “Studying the Song Development Process: Rationale and Methods.” *Annals of the New York Academy of Sciences* 1016(1): 348–63. http://doi.wiley.com/10.1196/annals.1298.031.

Tchernichovski, O. et al. 2000. “A Procedure for an Automated Measurement of Song Similarity.” *Animal Behaviour* 59(6): 1167–76. http://linkinghub.elsevier.com/retrieve/pii/S0003347299914161.

Theofanopoulou, C., C. Boeckx, and E.D. Jarvis. 2017. “A Hypothesis on a Role of Oxytocin in the Social Mechanisms of Speech and Vocal Learning.” *Proceedings of the Royal Society B: Biological Sciences* 284(1861).

Wang, Xiaoyue et al. 2017. “Speech-Specific Categorical Perception Deficit in Autism: An Event-Related Potential Study of Lexical Tone Processing in Mandarin-Speaking Children.” *Scientific Reports* 7(1): 43254. http://www.nature.com/articles/srep43254.

Warlaumont, Anne S., Jeffrey A. Richards, Jill Gilkerson, and D. Kimbrough Oller. 2014. “A Social Feedback Loop for Speech Development and Its Reduction in Autism.” *Psychological Science* 25(7): 1314–24. http://journals.sagepub.com/doi/10.1177/0956797614531023.

Zann, Richard A. 1996. *The Zebra Finch: A Synthesis of Field and Laboratory Studies.* Oxford Uni. New York. https://www.journals.uchicago.edu/doi/10.1086/420003.

Appendix

Chapter 3

Pilot study: testing the effect of intranasal administration of an oxytocin-receptor antagonist in adult zebra finch directed singing

Theofanopoulou C. [1,3], Boeckx C. [1,2] & Jarvis E.D. [3,4]

[1]Section of General Linguistics, University of Barcelona, Barcelona, Spain

[2] ICREA

[3]Department of Neurogenetics of Language, Rockefeller University, New York, USA

[4]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

Abstract

Social interactions are thought to enhance the motivation to learn and perform species-typical behaviors, like singing in songbirds. However, the neurobiological mechanisms connecting social motivation and singing remain to be elucidated. Using zebra finches (*Taeniopygia guttata*), a commonly studied model for vocal learning, we show that manipulations of the nonapeptide hormone oxytocin (OT) affects socially modulated singing. We administered an oxytocin antagonist intranasally in males, and then co-housed them with a female to elicit singing to her (directed-singing). We found that oxytocin-antagonist-treated males had a significant drop in the number of introductory notes in their directed love song, more similar to the levels found in undirected song without a female. To test whether the intranasal route of oxytocin crosses the blood-brain barrier in zebra finches, we assayed oxytocin receptor antagonist-SAP conjugate via immunohistochemistry (with a SAP-antibody) after intranasal versus intramuscular routes, relative to intracranial positive controls, and found that the product crossed the blood-brain barrier only with intranasal delivery. Given the convergent vocal learning behaviors and associated neural pathways in zebra finches and humans, our results imply that oxytocin modulates some social differences in production of learned vocalizations. This finding also has implications for intranasal administration of oxytocin that is currently used in patients with social deficits, albeit so far to study other aspects of cognition.

Introduction

Male zebra finches display two song behaviors: directed and undirected singing. Directed song is addressed almost exclusively to females, it is usually accompanied by a courtship dance, and it is preceded by more introductory notes and sung faster than undirected song (Sossinka and Böhner 1980; Jarvis et al. 1998) (Figure 1). Undirected song is performed when the male is in the presence of other males, alone, or outside a nest occupied by its mate. Males also produce less motifs per bout and more acoustic variability during undirected compared to directed song (Sakata, Hampton, and Brainard 2008)

Theofanopoulou, Boeckx, and Jarvis 2017 hypothesized that oxytocin (OT) may modulate or control social motivation for vocal learning as well as social differences in production of learned vocalizations, including directed versus undirected singing. In songbirds, experimental manipulation of the oxytocinergic system with OT agonist and antagonist have been made mostly in the context of pair-bonding and aggression, with very few and some controversial reports on how these treatments affected singing, probably due to different treatment sites (Goodson, Lindberg, and Johnson 2004; Klatt and Goodson 2013; Pedersen and Tomaszycki 2012). One study found that OT increases the amount of directed singing to females (Pedersen and Tomaszycki 2012), whereas another (Goodson, Lindberg, and Johnson 2004) found no effect in either directed or undirected singing. Klatt and Goodson 2013 did not find any significant result either, but they note that males treated with an oxytocin-antagonist (OTA) tend to sing more courtship songs. All these studies focused only on the amount of singing,

which can vary between birds, and did not study any other aspect of their song that has been linked directly to differences in directed and undirected singing.

Here we sought to test our hypothesis, by manipulating OT and measuring adult singing behavior. We found that intranasal administration of an OTA crossed the blood brain barrier and impacted courtship directed song of adult male zebra finches.

Methods

Animals

Animal care and experimental procedures were approved by the Institutional Animal Care and Use Committees of Duke University and Rockefeller University. We used male and female zebra finches from the breeding colonies of each University. Adult birds ranged in age from 90 days to 1 year.

Oxytocin antagonist administration and song behavior

The oxytocin antagonist (OTA; ²d(CH2)51, Tyr(Me)2,Thr4, Orn8, des-Gly-NH29]-Vasotocin trifluoroacetate salt; Bachem, catalog #4031338) was dissolved in water (10μg/ml) and then diluted in sterile saline (1μg/ml). 12 μl of the solution were administered to the birds intranasally (6 μl in each nostril) with a pipette tip. Saline was administered intranasally as a vehicle control (12 μl). Administration was done while holding the bird in the experimenter's hand (Figure 2).

Males were co-housed with females and song recordings started right after OTA or saline administration. Song recording was done with Sound Analysis Pro (Tchernichovski et al. 2004). The number of introductory notes was counted manually per song and per motif and it was averaged from a minimum number of 10 song bouts over 2 sessions per bird. A paired t-test was conducted, with standard error of the mean (SEM), T-values, Degrees of Freedom and p-values calculated for the average of the number of introductory notes sang in the saline versus the OTA group, further grouping the average number of introductory notes per song and per motif.

Assessing crossing of blood brain barrier

In order to assess whether an intranasal delivery of an OT conjugate is able to cross the Blood Brain Barrier (BBB)in zebra finches, we used an oxytocin saporin antagonist (Oxytocin-SAP [IT-46], Advanced Targeting Systems), which specifically destroys neurons that have oxytocinergic receptors on their surface (Baskin et al. 2010). Saporin is a plant enzyme with N-glycosidase activity that depurinates a specific nucleotide in the ribosomal RNA 28S, thus irreversibly blocking protein synthesis. We diluted OT-SAP in PBS (1 μg/ml) and injected 12 μl intranasally (6 μl in each nostril with a pipette tip), 6 μl intramuscularly and 2 μl intracranially (with a Hamilton syringe). We performed the injections to 2 birds for each route-group (intranasal, intramuscular, intracranial).

Ten minutes after the injection, the birds were perfused under deep Nembutal anesthesia with 60 ml of PBS, followed by 60 ml of 4% paraformaldehyde in PBS. Brains were removed and cryoprotected in 30% sucrose overnight at 4°C. Saggital sections (14 μm) were cut on a freezing sliding microtome. Sections were processed for immunohistochemistry using a polyclonal antibody against rabbit SAP (AB-41AP, Advanced Targeting Systems), which recognizes saporin.

Brain sections were fixed with 4% PFA in PBS for 15 min at 4°C, then permeabilized with 0.1-0.5% triton-x 100 for 10 min at –20°C. Nonspecific binding was blocked by 2 × 10 min washes

in PBS (pH 7.0) with 5% skim milk (PBSM) and 0.1-0.5% Triton X-100 for 1h at RT (500µl per slide). Sections were then incubated overnight at 4°C with a 1:1000 dilution of the SAP antibody in blocking solution (500µl per slide), followed by 3 x 5 min PBS washes at RT and incubation with the secondary anti-rabbit IgG (Sigma; St. Louis, MO) diluted in the blocking solution for 1 hr at RT. After 3 x 5 min washes with PBS at RT, we mounted the sections on glass slides covered slipped in Vectashield Antifade Mounting Medium with DAPI, and let them sit in the dark overnight, before taking the pictures with a microscope.

Results

In this pilot study we wanted to test a possible effect of an OTA on singing behavior of male zebra finches. We opted for the intranasal route of oxytocin (IN-OT) to make our study translational to the array of studies in humans where IN- OT appears to be a promising approach to treat several clinical symptoms, like social cognition in autism (Veening and Olivier 2013).

Blocking oxytocin impacts directed song

We found that OTA-treated males decreased significantly the number of introductory notes they sang per song bout (p=0,0175) and per motif (p=0,0032) compared to the song of the same males when they were treated with saline (Figure 3). This reduction in the number of introductory notes was similar to the number that the birds sing during undirected song in the absence of a female. This suggests that OT might be involved in the circuitry that drives the sociosexual motivation differences of the birds to sing female directed song and ends up affecting their singing performance.

Oxytocin antagonist crosses the blood brain barrier

It is possible that intranasal administration of OTA impacted song behavior by crossing the Blood-Brain-Barrier (BBB) and directly affecting the brain, or that it affected peripheral systems. Whether intranasal delivery crosses the BBB remains a highly debated issue: imaging studies show that IN-OT reaches the brain in humans and shows higher OT cerebrospinal fluid (CSF) concentrations in rodents, macaques and humans (reviewed in Quintana et al. 2018). But it remains unknown whether it reaches the brain directly or it sends afferent feedback to the brain from peripheral organs, and it has never been tested in birds.

To test whether an IN-OT delivery is able to cross the BBB in zebra finches, we delivered an OT-SAP conjugate by intracranial, intranasal or intramuscular routes, and measured whether the OT-conjugate was present in the brain via immunohistochemistry with a SAP-antibody. For our positive control intracranial delivery, we found unequivocal staining at the injection site, near the HVC song nucleus (Figure 4A). For the intranasal delivery, we found less robust, but still evident binding of the SAP-antibody (Figure 4B, staining in the ventricular system). For the intramuscular delivery, we did not find any stained cells (Figure 4C, is representative for the rest of the brain). These findings show that the oxytocin product crosses the BBB, but only with intranasal delivery and not intramuscular delivery.

Discussion

Our findings are consistent with the hypothesis that OT modulates social context differences in production of learned song. In the presence of an IN-OT antagonist, male zebra finches produced less introductory notes at levels similar to that normally seen during undirected song. The males still, however, directed their songs to the females. In contrast to what is predicted from (Klatt and Goodson 2013), we did not note an increase in amount of singing in OTA-treated males relative to saline control animals. This suggests that IN-OT may not so much

affect the ability to produce courtship song, but affect the specific sequence and acoustic features of courtship during directedsinging.

The finding that intranasally administered, but not intramuscular administered, OTA crosses the BBB opens up a venue for non-invasive surgeries in zebra finches, that can be translated in other species. Importantly, it lends support to the studies in humans and rodents showing functional associations after IN-OT or IN-OTA, rendering clearer that these associations are most likely due to direct modulation of neural function. This comes in line with studies showing elevated CSF and blood concentrations of OT following IN-OT in humans and in macaques (Dal Monte et al. 2014; Striepens et al. 2013) and with IN-OT-induced changes in resting regional cerebral blood flow in humans (Paloyelis et al. 2016).

This finding contributes to enhancing the validity and reliability of clinical trials investigating the therapeutic potential of IN-OT in humans. Additionally, we suggest that IN-OT could be a promising therapeutic method for speech deficits. There are already some studies pointing to the direction, where IN-OT in autistic patients leads to a more efficient and long-lasting performance in a speech comprehension task (Hollander et al. 2007; Pfundmair et al. 2016). Our findings suggest that future studies in humans with OT manipulations should also focus on speech production and speech learning.

As proposed in Theofanopoulou, Boeckx, and Jarvis 2017, the site of action of OT on song behavior could be made possible through oxytocinergic neurons from the hypothalamus that project directly to the RA and HVC song nuclei, as well as the VTA in the midbrain (Figure 5). From the VTA, OT can act synergistically with VTA's dopaminergic projections to AreaX (Lewis et al. 1981; Hara et al. 2007). DA levels in Area X are higher during directed singing (to females) than undirected singing, due to differential activity of the re-uptake transporter (a noradrenaline transporter in birds), in the VTA axons within Area X (Hara et al. 2007). In addition to that, unilateral lesions of the VTA dopaminergic projections reduce singing-driven Immediate Early Gene (IEG) expression in Area X in both contexts (Sasaki et al. 2006).

What remains to be done in our project is to increase the sample size and measure additional features of song, such as speed, syntax, and acoustic variability (work in progress). It would be interesting to include in the song analysis others features like number of motifs per bout and variability in sequence, duration and pitch. Overall, our preliminary findings are supportive of a role of oxytocin in at least modulating singing behavior in songbirds.

Figures

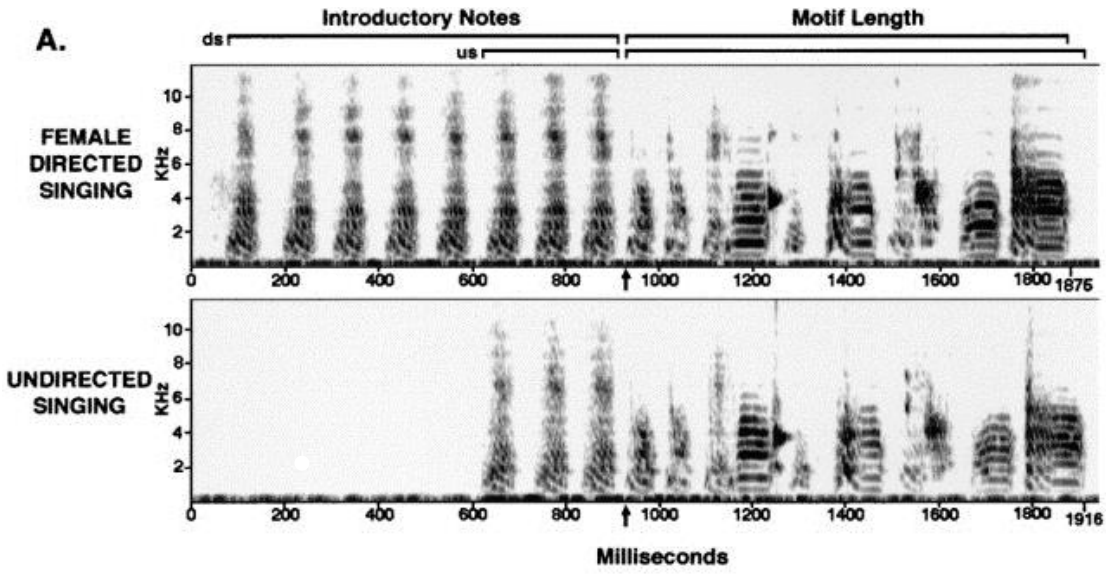


Figure 1: Sonograms from a male zebra finch when singing directed vs. undirected singing (right) (reproduced from Jarvis et al. 1998).

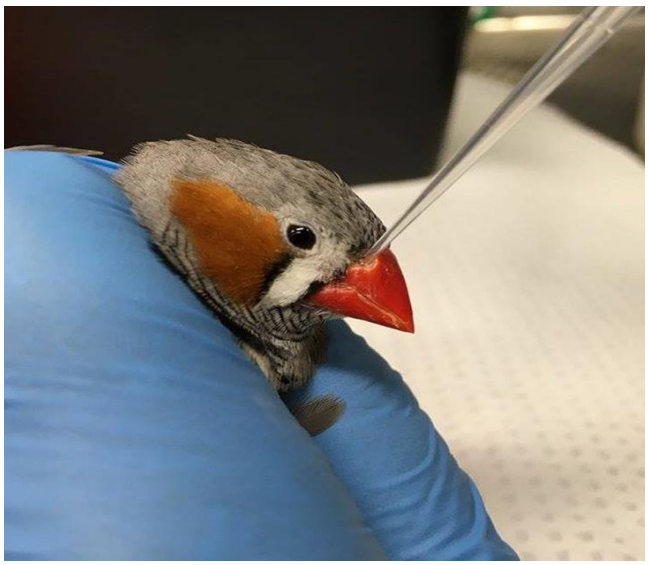


Figure 2: Demonstration of the intranasal treatment with a pipette tip.

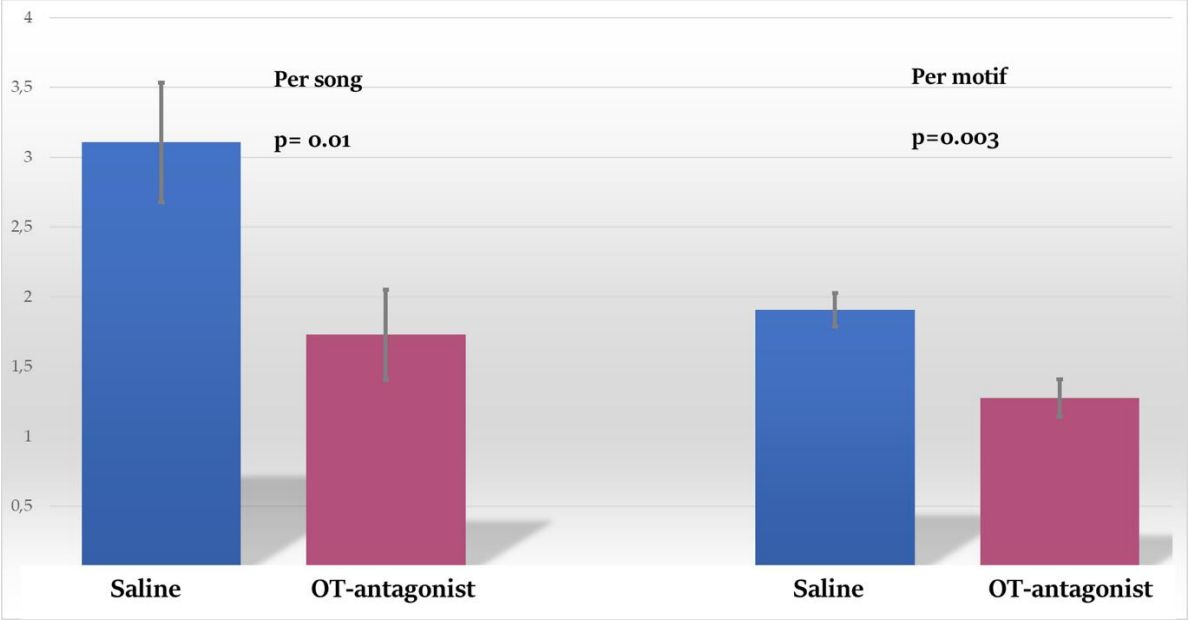


Figure 3: The number of introductory notes shown for each administration-group (Saline, OTA), averaged out by the number of female directed songs or motifs sang by male zebra finches. Error bars are SEM; p values are from a paired-t-test, n = 3males.

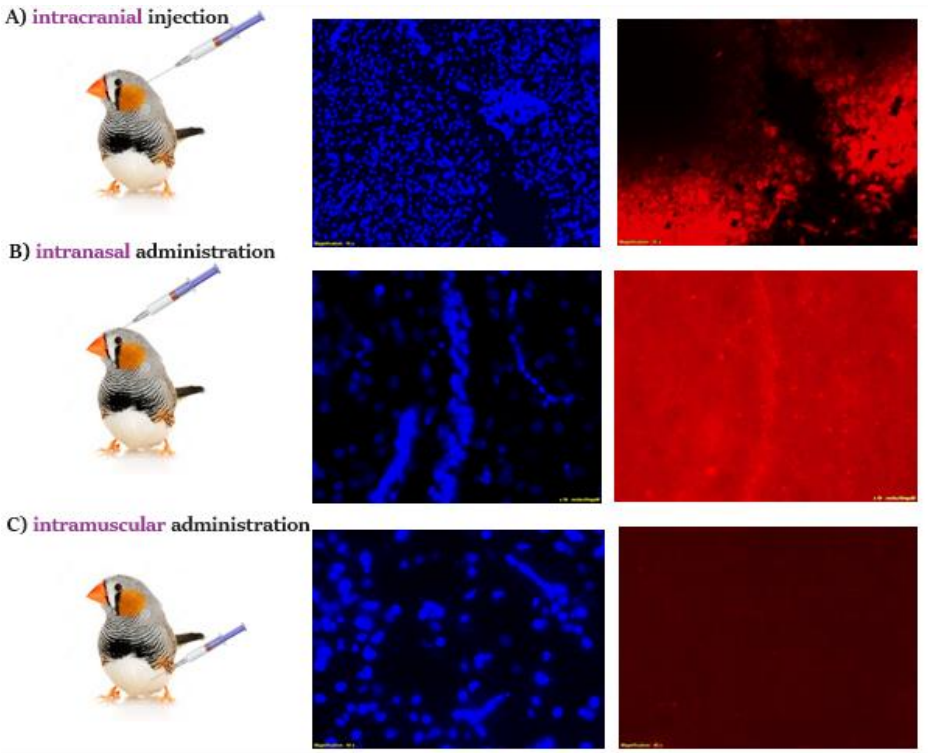


Figure 4: Test if OT crosses the blood brain barrier in zebra finches. For each delivery-group we show SAP expression (left-blue: DAPI, right-red: SAP). **A**) Intracranial, SAP expression at

injection site (near HVC). **B**) Intranasal, SAP expression in the ventricular system. **C**) Intramuscular, no SAP staining was found throughout the brain.

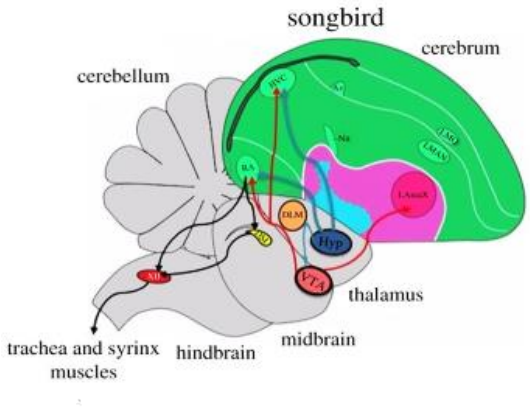


Figure 5: Proposed oxytocinergic (in blue) and dopaminergic (in red) projections into the vocal learning circuits. We propose oxytocinergic neurons from the Hyp project to RA, HVC and VTA; VTA makes a strong dopaminergic projection to AreaX and weaker ones to HVC and RA. Black arrows, connectivity of the proposed system with the brainstem. Abbreviations: HVC, HVC nucleus; LMAN, lateral magnocellular nucleus of anterior nidopallium; RA, robust nucleus of arcopallium; Area X, area X of the striatum; Hyp, hypothalamus; VTA, ventral tegmental area; DLM, dorsal lateral nucleus of the medial thalamus; Av, nucleus avalanche; LMO, lateral oval nucleus of the mesopallium; NIf, interfacial nucleus of the nidopallium; DM, dorsal medial nucleus of the midbrain; XII, 12th nucleus, tracheosyringeal part. (Adapted from Theofanopoulou, Boeckx, and Jarvis 2017)

References

Baskin, Denis G. et al. 2010. "A New Oxytocin-Saporin Cytotoxin for Lesioning Oxytocin-Receptive Neurons in the Rat Hindbrain." *Endocrinology* 151(9): 4207–13. https://academic.oup.com/endo/article-lookup/doi/10.1210/en.2010-0295.

Dal Monte, Olga et al. 2014. "CSF and Blood Oxytocin Concentration Changes Following Intranasal Delivery in Macaque" ed. David A. Slattery. *PLoS ONE* 9(8): e103677. https://dx.plos.org/10.1371/journal.pone.0103677.

Goodson, James L, Laura Lindberg, and Paul Johnson. 2004. "Effects of Central Vasotocin and Mesotocin Manipulations on Social Behavior in Male and Female Zebra Finches." *Hormones and Behavior* 45(2): 136–43. http://linkinghub.elsevier.com/retrieve/pii/S0018506X03002381.

Hara, Erina, Lubica Kubikova, Neal A. Hessler, and Erich D. Jarvis. 2007. "Role of the Midbrain Dopaminergic System in Modulation of Vocal Brain Activation by Social Context." *European Journal of Neuroscience* 25(11): 3406–16. http://doi.wiley.com/10.1111/j.1460-9568.2007.05600.x.

Hollander, Eric et al. 2007. "Oxytocin Increases Retention of Social Cognition in Autism." *Biological Psychiatry* 61(4): 498–503. http://linkinghub.elsevier.com/retrieve/pii/S0006322306007293.

Jarvis, Erich D et al. 1998. “For Whom The Bird Sings.” *Neuron* 21(4): 775–88. http://linkinghub.elsevier.com/retrieve/pii/S0896627300805942.

Klatt, James D., and James L. Goodson. 2013. “Oxytocin-like Receptors Mediate Pair Bonding in a Socially Monogamous Songbird.” *Proceedings of the Royal Society B: Biological Sciences* 280(1750).

Leblois, A., B. J. Wendel, and D. J. Perkel. 2010. “Striatal Dopamine Modulates Basal Ganglia Output and Regulates Social Context-Dependent Behavioral Variability through D1 Receptors.” *Journal of Neuroscience* 30(16): 5730–43. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5974-09.2010.

Lewis, James W., Susan M. Ryan, Arthur P. Arnold, and Larry L. Butcher. 1981. “Evidence for a Catecholaminergic Projection to Area X in the Zebra Finch.” *The Journal of Comparative Neurology* 196(2): 347–54. http://doi.wiley.com/10.1002/cne.901960212.

Paloyelis, Yannis et al. 2016. “A Spatiotemporal Profile of In Vivo Cerebral Blood Flow Changes Following Intranasal Oxytocin in Humans.” *Biological Psychiatry* 79(8): 693–705. https://linkinghub.elsevier.com/retrieve/pii/S0006322314007653.

Pedersen, A., and M. L. Tomaszycki. 2012. “Oxytocin Antagonist Treatments Alter the Formation of Pair Relationships in Zebra Finches of Both Sexes.” *Hormones and Behavior* 62(2): 113–19. http://dx.doi.org/10.1016/j.yhbeh.2012.05.009.

Pfundmair, Michaela, Franziska Lamprecht, Florentine M. von Wedemeyer, and Dieter Frey. 2016. “Your Word Is My Command: Oxytocin Facilitates the Understanding of Appeal in Verbal Communication.” *Psychoneuroendocrinology* 73: 63–66. https://linkinghub.elsevier.com/retrieve/pii/S0306453016304590.

Quintana, Daniel S, Knut T Smerud, Ole A Andreassen, and Per G Djupesland. 2018. “Evidence for Intranasal Oxytocin Delivery to the Brain: Recent Advances and Future Perspectives.” *Therapeutic Delivery* 9(7): 515–25. https://www.future-science.com/doi/10.4155/tde-2018-0002.

Sakata, Jon T., Cara M. Hampton, and Michael S. Brainard. 2008. “Social Modulation of Sequence and Syllable Variability in Adult Birdsong.” *Journal of Neurophysiology* 99(4): 1700–1711. http://www.physiology.org/doi/10.1152/jn.01296.2007.

Sasaki, A., Sotnikova, T. D., Gainetdinov, R. R., & Jarvis, E. D. 2006. “Social Context-Dependent Singing-Regulated Dopamine.” *Journal of Neuroscience* 26(35): 9010–14.

Sossinka, Roland, and Jörg Böhner. 1980. “Song Types in the Zebra Finch Poephila Guttata Castanotis 1.” *Zeitschrift für Tierpsychologie* 53(2): 123–32. http://doi.wiley.com/10.1111/j.1439-0310.1980.tb01044.x.

Striepens, Nadine et al. 2013. “Elevated Cerebrospinal Fluid and Blood Concentrations of Oxytocin Following Its Intranasal Administration in Humans.” *Scientific Reports* 3(1): 3440. http://www.nature.com/articles/srep03440.

TCHERNICHOVSKI, O et al. 2004. “Studying the Song Development Process: Rationale and Methods.” *Annals of the New York Academy of Sciences* 1016(1): 348–63. http://doi.wiley.com/10.1196/annals.1298.031.

Theofanopoulou, C., C. Boeckx, and E.D. Jarvis. 2017. “A Hypothesis on a Role of Oxytocin in the Social Mechanisms of Speech and Vocal Learning.” *Proceedings of the Royal Society B: Biological Sciences* 284(1861).

Veening, Jan G., and Berend Olivier. 2013. “Intranasal Administration of Oxytocin: Behavioral and Clinical Effects, a Review.” *Neuroscience & Biobehavioral Reviews* 37(8): 1445–65. https://linkinghub.elsevier.com/retrieve/pii/S0149763413001085.

Appendix

Chapter 4

A proposed universal nomenclature for the oxytocin and vasotocin ligand and receptor families and their evolutionary history

Constantina Theofanopoulou[1,2], Gregory Gedman[2], James A. Cahill[2], Cedric Boeckx[1,3] and Erich D. Jarvis[2,4]

[1]Section of General Linguistics, University of Barcelona, Barcelona, Spain
[2]Laboratory of Neurogenetics of Language, Rockefeller University, New York, USA
[3]ICREA
[4]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

Abstract

Oxytocin (OT) and vasopressin/vasotocin (VT) are important nervous system transmitter ligands that function through specific receptors to control a diverse set of brain functions, from social bonding to body homeostasis. Due to differential naming of the ligands by biochemists according to small amino acid differences between species and to high sequence identities between the oxytocin (OTR) and vasopressin/vasotocin receptors (VTR), there is often confusion about their orthology and paralogy, making it difficult to readily translate findings across species. Here we performed genome analyses across vertebrates to determine the evolution of these gene families and propose a revised and universal nomenclature for all vertebrates. This included identity (BLAT/BLAST) and synteny (GEvo, SynMap and SynFind) analyses on putative OT-VT ligands and their receptors in the genomes of 33 species that span all major vertebrate lineages, and newly re-sequenced species with long-read technology that filled in gaps and corrected errors in previous shorter-read assemblies (e.g. sea lamprey germline genome). Our findings confirm and present further evidence that OT and VT are adjacent paralogous genes that formed as a local genomic duplication event near the origin of vertebrates, with VT being the parental gene of OT. We propose that this duplication occurred via DNA transposable elements around the OT gene. What has been called mesotocin, isotocin, or oxytocin-like in non-mammalian species are all the same gene, namely oxytocin; vasotocin in all non-mammalian vertebrates is the same as vasopressin in mammals. Thus, following the standard practice in molecular biology and genomics, we propose that these two genes be given the same orthologous names across vertebrates and paralogous names relative to each other, namely oxytocin and vasotocin. We identified six OTR-VTR receptors among vertebrates, with some receptors absent in some clades. We traced the evolutionary history of these receptors, and propose that the OTR-VTR receptors arose from an ancestral invertebrate-like VTR that through a combination of large segmental and whole genome duplications, originally formed vertebrate VTR1 and VTR2, followed by three genes per family, VTR1 family (OTR, VTR1A, VTR1B) and VTR2 family (VTR2A, VTR2A, VTR2B), and then by one to two losses per vertebrate lineage. This is the first study to propose a universal nomenclature for the oxytocin and vasotocin ligand and receptor gene families. This new nomenclature should prevent further confusion and errors, allow easier translation of findings across vertebrates, and foster more informative design of functional experiments across species.

Introduction

Oxytocin (OT) and vasopressin/vasotocin (AVP/VT; henceforward referred to as vasotocin, VT in this study), depending on the brain region and release site, can act as hormones or neurotransmitters that function through their respective receptors to regulate a wide range of partially overlapping biological processes (Knobloch and Grinevich 2014; Meyer-Lindenberg et al. 2011). For oxytocin, this includes functioning in uterine contractions and pregnancy, milk-ejection, copulation and orgasm, attachment between parents and their young, bond formation, suppression of stress, thermoregulation, nesting behavior, olfactory processing, eye-contact, recognition of familiar individuals and social vocalizations among others. For vasotocin, this includes its well-known functions as an antidiuretic and regulator of blood pressure increase,

and lesser known but important functions in reproduction, pair-bonding, and parental care. Some of these functions are specific to individual lineages, such as nesting behavior in birds and milk-ejection in mammals.

OT and VT are closely related structurally/biochemically and evolutionarily. Biochemically, they show small amino acid differences across species, which led biochemists that discovered them in the pre-genomic era to give different names to the similar molecules in different lineages: such as mesotocin in birds, reptiles and frogs; isotocin in teleost fish; and valitocin in sharks, for the apparent oxytocin compliment of mammals; and vasopressin in mammals for the apparent vasotocin compliment in other vertebrate lineages (Acher and Chauvet 1988). This terminology was then applied to their respective receptors, such as vasopressin receptor (AVPR or VPR) in mammals and vasotocin receptor (VTR) in birds.

Evolutionarily, it has been hypothesized that OT and VT are the product of a local duplication event that took place before the origin of vertebrates (Hoyle 1999). Concerning evolution and homology of the receptors, there are many proposals in the literature (Daza et al.2012; Yamaguchi et al. 2012; Lagman et al. 2013; Mayasich and Clarke 2016). Two recent broad genome-wide views are: one by Lagman et al. 2013 who proposed OT and VT receptors evolved through two rounds (2R) of whole genome duplications (WGD) in the cyclostome ancestor; and one by Mayasich and Clarke 2016 who proposed as an alternative scenario that they evolved by a shared agnathan-gnathostome 1R of WGD plus segmental duplications. However, these studies used incompletely assembled genes and genomes, with some genes missing or misplaced, and thus could not resolve these or other hypotheses.

As a result, the varied biochemical-based lineage terminology and various evolutionary-based terminologies has led to confusion on the orthology and paralogy of OT and VT, and their receptors. Different investigators use different terminologies for the same gene, some with the belief that orthologous genes are not the same gene and others with the belief that some receptors are the same gene (Parry et al. 2000; Gubrij et al. 2005; Baeyens and Cornett 2006; Yamashita and Kitano 2013), when they are not. This in turn makes it difficult for scientists to readily translate findings across species, such as in our and others' studies of wanting to translate findings on OT, VT, and their receptors between distantly related vocal learning species, like humans and songbirds (Leung et al. 2011; Theofanopoulou et al. 2017; Ondrasek et al. 2018).

In this study, we aimed to clear up this problematic nomenclature. Doing so required gaining a better understanding of the evolutionary history of the OT and VT ligands and their receptors. We analyzed the genomes of 33 species that span all major vertebrate lineages and 4 outgroup invertebrate lineages (Table 1). These included newly re-sequenced species (e.g. inshore hagfish (*Eptatretus burgeri*) and species with long-read technology that filled in gaps and corrected errors in previous shorter-read assemblies (e.g. sea lamprey (*Petromyzon marinus*)) (Smith et al. 2018). We based gene homology not only on sequence identity and phylogenetic gene trees, but also on conserved synteny of the genomic territory around the genes of interest. Synteny is one of the most reliable criteria for establishing gene orthology, where genes that are positioned near or next to each other on the genome in one species are likely to be found close to each other on a single chromosome/contig in another species (Ghiurcuta and Moret 2014). We show that the common practice of basing gene orthology on sequence identity only for this gene family is not sufficiently reliable, as some orthologous relationships were obscured by sequence divergence or by different divergence rates in different lineages, leading to incorrect gene nomenclatures. We also analyzed transposable elements to search for possible mechanisms to explain gene duplications and translocations.

We propose that OT and VT are paralogous genes that arose through a local duplication via transposable elements at the origin of vertebrates. We propose that the OT and VT receptors evolved by a combination of segmental and WGD events, leading to six receptors in total near the origin of vertebrates, with losses and additions of some types in subsequent specific

vertebrate lineages. With a better understanding of their orthology and evolutionary history, we also now propose a universal nomenclature for these genes across all vertebrates (Table 2).

Results

We collected and analyzed available genomes of 33 species spanning all major vertebrate clades, and 4 invertebrate outgroups, including species representing the most basal branches (Table 1). We also collected the available literature on the terminology and orthology diversity of oxytocin and vasotocin ligands and their receptors, to make sure we captured all claims of the presence versus absence in specific species and the nomenclatures used (Table 2). Then in all the genomes collected, we searched for oxytocin and vasotocin genes and their receptors, using pair-wise BLAST analyses of sequences from each major vertebrate class. We then used whole genome alignments across species and multi-chromosome alignments within species, to analyze synteny from microchromosomal to macrochromosal scales. We then performed phylogenetic tree inference on gene and gene-family trees, and assessed congruence between synteny, sequence identity, and inferred phylogeny based on sequence identity.

For the synteny analyses, we only concluded that a gene was deleted (loss) if we could find genes surrounding the loss region without a genome assembly gap. We also used the most complete assembly of the species analyzed, including several we and others generated via the G10K Vertebrate Genomes Project (VGP; https://vertebrategenomesproject.org) with long read technology where there were fewer gaps and more complete and accurate assemblies (Table 1). These include the recently re-assembled hummingbird (*Calypte anna*) and zebra finch (*Taeniopygia guttata*) avian genomes (Korlach et al. 2017), a sea lamprey (*Petromyzon marinus*) germline genome (Smith et al. 2018), and a hagfish (*Eptatretus burgeri*) genome (Eburgeri_3.2, released in June 2018), as representative of the only extant lineages of jawless vertebrates (agnatha). In the majority of the genomes, we found new orthologs of one or more genes or resolved previous tentatively identified orthologs. Below we present findings and revised nomenclature first on the ligands, and then on the receptors, which we find have a more diverse evolutionary history.

Syntenic vasotocin and oxytocin orthologs are found in nearly all vertebrate species. Based on BLAST searches, sequence identity, and local synteny analyses within a 10-gene window, we found the ortholog of the human vasotocin (VT) gene (e.g. arginine vasopressin) in all vertebrate lineages analyzed (Tables 3, S1a; Figure 1). In contrast, we found the oxytocin (OT) ortholog in nearly all lineages, with the exceptions of lampreys and hagfish (Table 3, S1b, S5; Figure 10). Determination of absence of OT required ruling out that the one gene found in lamprey and hagfish were the ortholog of VT and not OT. A previous study could also not find the OT ortholog in the Japanese lamprey (*Lethenteron japonicum*), but had uncertainty due to the assembly being from the sized down and rearranged somatic genome for this species (Gwee et al. 2009). Using the first assembly of a lamprey germline genome (Smith et al. 2018), we find robust evidence that lampreys possess only the VT gene. For the hagfish genomic territory of the putative VT we found only 6 genes in the territory; these genes were not annotated, but we did so with the 'Region Comparison' tool in Ensembl, against the human, zebrafish and lamprey genomes (Table S5). None of these genes were encountered in the neighboring territory of lamprey VT; but in the 'Gene Tree' (ENSGT00390000004511) available for this gene, we found that it formed an immediate node with lamprey VT, and not vertebrate OT. Thus, we believe this to be more of the homolog of VT than OT. We could not find any other orthologs in any of the vertebrate genomes.

With regards to invertebrates, our analyses of 4 species (Table 1) supports the findings of Gwee et al. 2009, who identified one single gene in most invertebrates. In amphioxus (*Branchiostoma floridae*), the structure and sequence identity were more similar to what we designated as VT in lampreys and the VT relative to OT in all other vertebrates. Gwee et al.

2009 reviewed the presence of only one homolog in other invertebrates studied thus far, with all homologs forming outgroups of the lamprey VT in the phylogeny. But the synteny of genes in amphioxus was not conserved in other invertebrates (specifically in *Ciona intestinalis*), and the synteny of genes in vertebrates was not conserved either in amphioxus or in *Ciona* (Gwee et al. 2009). Lastly, Liutkeviciute et al. 2016 found only one OT/VT-like hormone in the vast majority of the insect species they studied (229 out of 233).

In nearly all vertebrate species, both OT and VT were present on the same chromosome and were directly syntenic with no other genes in-between (Tables S1a, b). The exceptions were the teleost fishes, where the genes were separated by four or more other genes, sill in the same syntenic locus, indicative of a local translocation or inversion; further in zebrafish (*Danio rerio*) they were on separate chromosomes (also noted by Gwee et al. 2009). In the spotted gar (*Lepisosteus oculatus*), the synteny followed the pattern of all the other vertebrates, indicating that of the ray-finned fishes, only the teleost (and not the holostei, like the spotted gar) display rearrangements relative to the ancestor. This is consistent with studies showing that teleost fish genomes have experienced a high rate of inter/intrachromosomal rearrangements, most likely due to the teleost-specific whole genome duplication (Jaillon et al. 2004).

Oxytocin appears to have evolved from vasotocin by a local gene duplication event. The local synteny of VT and OT, and the absence of an OT ortholog in species representing the basal branches of vertebrates as well as invertebrates, suggest that VT could have given rise to OT. Gwee et al. 2009 had also hypothesized that the ancestral gene was VT, based on the higher homology between the gene they found in the Japanese lamprey and a VT-like gene found in the invertebrate amphioxus, and secondly based on the orientation of the genes: they found that OT and VT were located on the same strand of DNA, in a tail-to-head orientation (same direction) in all vertebrates, apart from placental mammals (tail-to-tail orientation, with the OT inverted relative to VT). Based on that the genes are not inverted in the opossum (*Chironectes minimus*) (a marsupial mammal), they assumed the inversion took place in the origin of placental mammals. We confirmed that the orientation is inverted in all placental mammals we studied, but also in the platypus (*Ornithorhynchus anatinus*) (a monotreme mammal) and the kangaroo rat (*Dipodomys ordii*) (a marsupial) (Table S2). Considering this additional data, we propose that the OT inversion occurred at the origin of mammals, and that the tail-to-head orientation found in opossum is a lineage specific re-inversion, likely due to the local duplication of both the VT and the OT regions in this species. We also identified that OT was inverted relative to VT in the spotted gar as well, indicative of an inversion specific to holostei or to the spotted gar. It seems that OT underwent inversions relative to VT two independent times in vertebrates.

We further tested the hypothesis of OT being a tandem duplication of VT searching for evidence beyond orientation and homology, namely for DNA transposal elements (TE). We found TEs (Figures 2, S1) around the OT gene (in human (*Homo Sapiens*) and chimpanzee (*Pan troglodytes*)), but not around the VT gene; these TEs had terminal inverted repeats (TIR) sequences that are known to transpose through a cut-and-paste mechanism creating an extra copy at the donor site (Wicker et al. 2007). We also searched for further features that have been independently encountered in the majority of duplicated genes, like intron shortening and increase in intronic GC-content (Rayko, Jabbari, and Bernardi 2006). Alignment of the introns in the human OT and VT, revealed that both OT-introns were shorter than VT-introns, with the first intron of OT being also 13% richer in GC-content (77.9% vs. 64.6%). The human OT gene body was also enriched in Alu elements (14,77%) compared to 0% in the VT gene body, a characteristic feature of duplicated regions (Bailey, Liu, and Eichler 2003). Although DNA TEs were not annotated in the elephant shark (*Callorhinchus milii*), we found a strong decrease in the length of only the 1st OT intron compared to the 1st VT intron (3226 bp vs 1158 bp), but their GC-content was the same. The totality of the findings of synteny, presence versus absence

of genes, inversions, TEs, Alus, intron lengths, and GC content, all suggest that the OT gene is a locally duplicated copy of the VT ancestral gene (Figure 2) followed by some greater divergence of these introns and GC content in different lineages.

A universal nomenclature for oxytocin and vasotocin genes. Based on these findings and the preponderance of findings in the literature, we propose a unified nomenclature where the names oxytocin (OT) and vasotocin (VT) are used for these genes present in all jawed vertebrates, and VT in all vertebrates and closely related invertebrates. We believe that they should be named in a way that portrays their evolutionary history, as is standard practice with other genes in the genomes that are orthologous across species and paralogous within species. Based on standard practice in molecular biology, paralogous genes are given the same root-name (e.g. FOXP) ending with different numbers (FOXP1, FOXP2, FOXP3, FOXP4, etc.). According to this practice, these two peptides would be named vasopressin1 (AVP1) and vasopressin2 (AVP2), vasotocin1 (VT1) and vasotocin2 (VT2), or oxytocin1 (OT1) and oxytocin2 (OT2). Since we realize that this would be a big shift from the already existing nomenclature, we propose that the common origin of these genes be portrayed through the shared ending name –tocin, and paralogy conveyed through different root names oxy- and vaso-. Vasotocin is a name that is already being used by most non-mammalian scientific communities (Table 2). Even within the mammalian scientific community there is no consensus on a gene name, since the name arginine vasopressin (AVP) entails that this gene has an arginine as the eighth amino acid, which is not the case for all mammals (such as in the guinea pig or the peruvian mouse, where there is lysine in this position; Acher and Chauvet 1988). For non-mammalian species, this means that what is now called mesotocin, isotocin, glumitocin, valitocin, aspargtocin, and neurophysin would be called by one orthologous name, oxytocin.

Six syntenic oxytocin and vasotocin receptors among vertebrates. Using identified sequences in at least one well-assembled genome of each major vertebrate lineage (e.g. human for mammals, chicken or zebra finch for birds, and germline sea lamprey genome for lampreys, etc.) we BLAST searched for oxytocin and vasotocin receptor hits in all other species and performed synteny analyses. Our microsynteny analysis within a 10-gene syntenic window revealed that there were six orthologous receptor types that we could confidently distinguish as separate paralogous genes among vertebrates (Figure 1; Tables 2, 3, S1c-h). However, most vertebrate species had different combinations of four of the six receptors, and fish had five in different combinations (Table 3), but none had all of them, indicating differences in gains and losses among vertebrate lineages. For greater clarity, to explain our results, we use the nomenclature of the root names for the ligands we propose above and evolutionary relationship-based names for the ending in analyses shown below. Using different names for different lineages makes explaining the findings on the receptors more cumbersome to follow.

The oxytocin receptor (OTR) ortholog was found in a relatively well conserved syntenic region in all vertebrate species examined (Tables 3, S1c). The same was the case for the gene commonly named arginine vasopressin receptor 1A (AVPR1A), named here vasotocin receptor 1A (VTR1A; Tables 3, S1d). In contrast, the vasotocin receptor 1B (VTR1B) ortholog was absent in lampreys and most fish, except coelacanth (*Latimeria chalumnae*) and elephant shark (*Callorhinchus milii*) (Figures 1, Table 3). We found the syntenic territory of the region in most fish, but not in the lampreys (Table S1e), indicating a possible gain in a vertebrate ancestor post-divergence with lampreys, followed by a loss in the teleost fish post-divergence with coelacanths and sharks.

VTR2A was not found in lampreys and birds (Tables 3, S1f). We did not find a similar microsyntenic 10-gene syntenic block in any of these species, but in birds, we noted that VTR2A is part of a larger block of ~20 genes that was deleted (Lovell et al. 2014), indicating a loss in birds. VTR2B was detected only in lampreys and fishes (Tables 3, S1g), but its syntenic territory was detected in all other vertebrate species, indicating a loss in the common ancestor of tetrapods (amphibians, reptiles, birds, and mammals). VTR2C was absent in some fish

species (coelacanth, medaka, tilapia, zebrafish) and all mammals, but its syntenic gene territory was present in all these species (Tables 3, S1h), indicating two independent losses.

For the syntenic territories of VTR2B and VTR2C we noted a complex pattern of apparent losses. In most species where VTR2B is present, VTR2B and OTR are located on the same chromosome. Using the human genome as a representative tetrapod, for genes we found in the territory of spotted gar VTR2B, those 5' to VTRB are located in two syntenic windows (3-5 Mb and 49-51 Mb) that are also 5 Mb before and 40 Mb after the location of human OTR respectively on the same chromosome 3 (Figures S14 and S15). This suggests that before the loss of VTR2B in the tetrapod ancestor, OTR and VTR2B were located on the same chromosome, as we find them in lampreys, fishes and coelacanths. Similarly, we searched for genes in non-mammalian (e.g. chicken (*Gallus gallus*)) VTR2C's syntenic block in humans, and found a region of highly conserved synteny on chicken chromosome 1 before VTR2C with human chromosome 7 (100-115 Mb) (Figure S16) and after VTR2C with human chromosome 12 (40-43 Mb) (Figure S17), corroborating similar evidence shown in (Lagman et al. 2013; Yamaguchi et al. 2012).

Tracing the synteny of VTR2C in sea lamprey was cumbersome, since the putative VTR2C (according to Mayasich and Clarke 2016) that is located on the same scaffold (10) with VTR1A does not share the microsynteny pattern we found in all other vertebrate species (Figure 3). We only found two genes (LRRN1, IMMP2L) that appear in synteny with VTR2C in some vertebrate species. When we compared genic and non-genic regions around the spotted gar and chicken VTR2C against the sea lamprey genome using SynFind, both lamprey PMZ_0042163-RA (scaffold 10) and PMZ_0045207-RA (scaffold 49) genes came up with equal syntenic depth to chicken VTR2C, while lamprey PMZ_0045207-RA was the first hit for spotted gar VTR2C (Figure S2). But for PMZ_0045207-RA (scaffold 49) the syntenic depth was higher only because of the larger territory around the receptor and not the immediate region of the receptor itself. As we show in Figure 3, microsynteny around lamprey PMZ_0045207-RA is highly reminiscent of the one we encountered around VTR2C in other vertebrate species. (Below we propose that this locus is really a duplication of OTR in lamprey).

In zebrafish we identified 2 copies each of OTR, VTR1A, and VTR2A, and 1 copy of VTR2B, and VTR2C, but with syntenic regions where the other copies of the latter two would be expected according to alignments with other vertebrate species (Table S3). The two copies and the apparent missing copies were all on the same chromosomes each, except for the two VTR1A copies. For example, genes that share identity with zebrafish VTR2B were on the same chromosome as the territory of an apparent deleted copy of zebrafish VTR2B. We surmised that these additional copies and deletions are due to large scale genome or chromosome segmental duplications in teleost, followed by losses of some of them. Based on synteny with other vertebrate lineages, we named the copy that had the most synteny to other vertebrates with 'a' and the one with less synteny with 'b', for example OTRa and OTRb, and VTR2Aa and VTR2Ab (Table S3).

We also identified an additional putative OTR orVTR receptor on lamprey scaffold 49 (PMZ_0014716-RA) very close to OT/VTR -49, never mentioned before in any other study that we are aware of (GL479461 and KE994228 scaffolds respectively in the somatic genome end a few genes before the new gene we identified). We did not find any match for this gene in any other vertebrate using SynFind, but we found its territory almost unaltered in elephant shark (on scaffold NW_006890088) but without the receptor gene. Next to this gene we found TRNT1 and LRRN1, which are the genes neighboring VTR2B in several vertebrate species (Table S1g). Thus, this gene is possibly a lamprey-specific duplication. We also found a species-specific duplication of a VTR in spotted gar (ENSLOCG00000000052, annotated as V2 receptor-like).

Lastly in hagfish we found two receptors in two different scaffolds, where the surrounding genes of one of them (ENSEBUG00000001467) are reminiscent of the OT/VTR-49-surrounding genes on scaffolds 49 and some genes on scaffold 27 in sea lamprey. The territory of the second receptor (ENSEBUG00000007964) did not match any known receptor in particular, but its surrounding genes were found scattered on sea lamprey scaffold 10 (Table S4).

Macrosynteny analysis: 40-100 gene window synteny analysis supports the proposed OTR-VTR orthology. We next assessed how far out the synteny we found holds, by increasing the gene window size until synteny became weaker, using the higher quality genome assemblies. Based on a 40 to 100 gene window synteny analysis, we found that VTR1A shows well conserved synteny when it comes to the following comparisons: coelacanth vs elephant shark and chicken, elephant shark vs gar and human, human vs frog and chicken (Figures 4b, S3d,e,i). In some species, there was synteny overlap with VTR2C as the next the best hit, something expected due to the very close positioning of VTR1A and VTR2C on the same chromosome in some species (frog, chicken, sea lamprey) that makes a 40-100 gene window mostly overlapping. OTR was found in clear syntenies with all species included in the analysis (Figures 4a, S3c). There is one case where chicken OTR appears to be more syntenic to gar VTR2B (not shown), but this is expected, since these two genes also lie very close to each other in the same chromosome in these species.

VTR1B appears to be clearly syntenic in the following species: human with coelacanth, frog and chicken, elephant shark with coelacanth, chicken with coelacanth and frog (Figure S3a). In the species where VTR1B is not present (sea lamprey, gar), it maps better to VTR1A. This concurs with VTR1B being the second best syntenic hit for VTR1A in several comparisons (coelacanth vs elephant shark and chicken, sea lamprey vs gar) and its high identity with VTR1A in most species (Table S6).

Studying the synteny of VTR2A was more complicated, since it is lost in most species and in the gar assembly, it appears as the only gene on a contig. Still human VTR2A mapped best to coelacanth VTR2A, as it is positioned between shark and coelacanth (Figure S3b). In the cases where VTR2A is lost, it showed more hits to VTR2C (e.g. coelacanth vs chicken). VTR2B appears in syntenies in the species that carry this gene (sea lamprey-gar, elephant shark-gar), while only lamprey VTR2B shares more synteny with elephant shark VTR1B than with VTR2B (Figure S3h). Lastly, VTR2C's synteny is robust in gar vs lamprey and elephant shark, human vs elephant shark, and chicken vs frog and gar. Interestingly, chicken VTR2C presents more synteny to lamprey OT/VTR-49, rather than to lamprey VTR2C (Figure S3f,g,j) according to other species' alignments.

Evolutionary history of OTR-VTRs, intraspecies macrosynteny. We next wanted to decipher the relationships among the different receptors and their evolutionary history. We surmised that if they evolved by 2 rounds of whole genome duplication, then we should find greater intra-species synteny in the more recent duplications versus the older ones. Thus, using the best assembled genome, human, we analyzed synteny in a 10 Mb window between chromosomes containing of all 6 receptors (those present and around the genes of those deleted), following an approach of Lagman et al. 2013, but focusing on syntenic gene families in a more strictly defined window. We found genes from the same gene families in syntenic blocks around all human OTR-VTRs, including those that survived in the territory of the deleted VTR2B and VTR2C (Figure 8, Tables S7, S8). However, only the SRGAP-gene family was present in the territory of all four human OTR-VTRs. Some gene families were shared in the surrounding regions of OTR/VTR1s and VTR2s, suggestive of their common origin (e.g. CNTN, LRRN, IRAK, L1CAM). Despite this clear intrachromosomal synteny, we do not detect any OTR-VTRs sharing significantly more synteny between them than others.

We made similar intra-species synteny and sequence identity comparisons within exons and introns of all the possible combinations of the sea lamprey OTR-VTRs. Our hypothesis was that if OTR got duplicated from VTR1A, for example, then intron and exon synteny, identity, and losses could be inferred, such as intron shortening in the younger gene. But, although we could find sequence identities between the exons and introns of all genes, again, no one gene showed consistently more similarity across introns/exons of another to make conclusions about gene family evolution (Figures S12-13). These findings support the hypothesis, that unlike the ligands, the receptors evolved by large regional if not chromosomal duplications, but either they occurred within a short period of each other leaving little if any synteny signature of the timing or that gene (and exon/intron) losses after the duplications were sufficiently random to tell the history of the timing of their duplications. In either case, they

would be followed by comparable numbers of different genes losses surrounding the OTR-VTR genes.

Phylogenetic analyses combined with syntenic orthology reveals receptor evolution. Next, we generated phylogenetic trees of the entire receptor gene family across vertebrate lineages using sequence alignments. We used the names of the genes we defined using synteny above in the resultant trees, replacing the diverse terminologies in the literature for each species. We used the Phylogenetic Maximum Likelihood method on both the exonic nucleotide sequences and the protein sequences (for the latter using the Ensembl tools). We only included sequences of exons that combined to more than 100 nucleotides (i.e more completely sequenced and assembled genes), as shorter sequences are known to give less reliable gene trees (Jarvis et al. 2014).

We found that topology of the resultant trees strongly supported our synteny-orthology receptor designations, with few exceptions (Figures 11 and 12). In most cases, the genes clustered according to orthology first, and then known species relationships second, supporting the hypothesis that the receptors diverged during early vertebrate evolution. In terms of specific evolution events, the trees provided strong support for OTR-VTR1 and VTR2 as separate gene subfamilies, consistent with previous analyses (Yamaguchi et al. 2012; Lagman et al. 2013; Mayasich and Clarke 2016). The trees (Figures 11 and 12) in conjunction with our synteny analyses (Table 3) suggest that there was a single VTR gene in an invertebrate ancestor of vertebrates (i.e. represented in *Ciona*). This receptor appears to have then duplicated at the origin of vertebrates in what we designate ancestral VTR1 and VTR2 receptors (Figures 11 and 12, branches labeled). Afterwards, the VTR1 receptor gave rise to OTR and VTR1A also near the origin of vertebrates, and then VTR1A gave rise to VTR1B post-divergence with lampreys (although the support for OTR is only 53% in the exon tree, this branching agrees with all the previously published phylogenies and is present in both trees). VTR2 gave rise to VTR2B and VTR2C, and later it gave rise to VTR2A post-divergence with lampreys (based on the exon tree, consistent with absence in lamprey). We think that this first duplication of VTR2 into VTR2B and VTR2C occurred prior to the divergence of bony fishes (in line with Mayasich and Clarke 2016), since we found VTR2A in teleost fishes, spotted gar and coelacanth. The topology of the protein suggests an alternative view, a split between VTR2A/B and VTR2C, but the bootstrap values are very low for this VTR2A clade topology (Figure 12), the VTR2B/C and VTR2A split with much higher bootstrap supports (Figure 11).

The phylogeny indicates that the three receptors in amphioxus are all lineage-specific duplications after divergence with vertebrates (Figure 11). The phylogeny supports other lineage-specific gains and losses within vertebrates suggested by the synteny analyses, including the teleost specific duplications of OTR, VTR1A, VTR2A loss of VTR2Ba in zebrafish, loss of VTR2A in birds, and loss of VTR2B in the ancestor of all tetrapods (amphibians, reptiles, birds, and mammals), and three independent losses of VTR2C in coelacanth, some teleost fish, and mammals (Table 3).

Differences in trees are that in the exon tree the sea lamprey VTR2B and VTR2C cluster as outgroups of the VTR2B/C clade, while in the protein tree they branch as outgroups of the teleost-VTR2C sequences (Figures 11 vs 12). Similar to that mentioned above, the exon tree topology is more consistent with the synteny. In the exon tree, sea lamprey VTR1A, OTR and OT/VTR-49 cluster as outgroups of the VTR1A/B clade, whereas in the protein tree VTR1A clusters as outgroup to VTR1A/B of other vertebrates. In both trees, lamprey OTR and OT/VTR-49 cluster together with strong bootstrap support, indicating they are maybe the result of a lamprey-specific duplication, and thus could be named OTRa and OTRb in lamprey. In the protein tree, hagfish ENSEBUG00000001467 clusters as an outgroup to VTR1A and VTR1B, and ENSEBUG00000007964 cluster as outgroup to spotted gar and tetrapod VTR2C. This clustering of hagfish and lamprey as outgroups is probably an artifact because of the high GC-content of the sea lamprey and hagfish genomes, which affects the codon use pattern and amino acid composition of protein-coding sequences and results in their paralogues being more identical to each other than to their orthologs in gnathostomes (noted also in Zhang et al. 2017

who used hagfish and lamprey sequences in their phylogeny). In either case, both support OTR as belonging to the VTR1 family.

Further, within specific receptor gene clades, not all species relationships break down with known relationships. For example, in the exonic tree, mammalian OTR appears as an outgroup to all other OTRs. These are common artifacts, due to several possible reasons: higher quality genomes in mammals, lineage-specific multiple mutations at a site, or divergent sequence length. For other sequences, where it was difficult to determine synteny, we found that the first exon of hagfish ENSEBUG00000007964 matched the full sequence region of the first exon of lamprey VTR2C (75%), while the second exon of hagfish ENSEBUG00000001467 was homologous to the second exon of lamprey VTR1A (67%).

Ancestral analyses support a single chromosome origin of vertebrate OTR-VTRs. Consistent with our phylogenetic analyses, our comparison of the chromosome fragments containing OTR-VTRs against the proposed Vertebrate-Ancestor-Chromosomes (VACs) in (Nakatani et al. 2007) suggest that OTR, VTR1A, VTR1B, VTR2B and VTR2C all date back to the same VAC-chromosome D; but the region of VTR2A (and a ~2Mb region surrounding it) had ambiguous synteny and was not included in the reconstruction (Table S9). Yun et al. 2015 suggested that, according to the Nakatani et al. 2007 paradigm, VTR2A maps back to VAC-chromosome F, unlike the rest of the OTR-VTRs that map to VAC (D). This, we believe, is an inaccuracy probably because the authors did not take into account that the actual region of VTR2A was missing from the reconstruction. Following the reconstruction of the chordate ancestor karyotype in Putnam et al. 2008, which was based on a higher quality amphioxus genome (intermediate length Sanger reads), we found instead that all OTR-VTRs, including VTR2A, originate back to the 13th putative ancestral chordate linkage group (Table S9). Following the presumptive ancestral chromosomes that Smith and Keinath 2015 and Smith et al. 2018 reconstructed based on high-quality sea lamprey somatic and germline genomes, respectively, we once again found that all our receptors correspond to the same ancestral chromosome in the vertebrate ancestor (AncD; Table S9). These findings suggest that ancestral linkage group #13 could be the original chromosome with the VTR gene passed onto and shared with invertebrates.

Regarding the number of ancestral OTR-VTRs in invertebrates, again consistent with the phylogenetic analyses we and others found that the vast majority of invertebrates have only one VTR and thereby OTR (e.g. *Tribolium castaneum*), while there are also some that carry more than one (e.g. 3 in *Branchiostoma floridae*, 2 in *Caenorhabditis elegans* and in *Ciona intestinalis*). From the 269 insect species studied in Liutkeviciute et al. 2016, 233 had one OTR-VTR, 7 more than one, and the rest none (our quantification based on their Table S1). Where invertebrates were included in both our protein (*C. intestinalis* and *C. elegans*) and exon (*Branchiostoma floridae*) trees, their OTR-VTRs cluster within these species, which is suggestive of lineage-specific duplications, instead of more than one progenitor gene that was inherited with vertebrates.

Sequence identity alone not sufficient to reveal gene orthology and evolutionary history When we aligned all sea lamprey, chicken and human OTR-VTRs against each other but also against spotted gar, coelacanth and elephant shark OTR-VTRs (Table S6), orthologous genes defined by synteny were not always the ones with the highest max scores or identities in many cases. Namely, it is not the case that the sequence of sea lamprey OTR will be more identical to the sequence of OTR in other species (e.g. coelacanth OTR) compared to other sequences (e.g. coelacanth VTR1A), nor is it the case that it will be more identical to a different sequence (e.g. VTR1A) in the rest of the species altogether.

We further wanted to test if such alignments can reveal the evolutionary history of some genes, particularly where VTR1B and VTR2A got duplicated from. For this reason, we aligned elephant shark VTR1B against all sea lamprey OTR-VTRs and compared introns and exons separately between VTR1B and the two sea lamprey OTR-VTRs with the highest max scores (OTR and VTR1A) (Figure S10). The identities and max scores were too similar to reveal putative ancestry. But in both our phylogenies (and all the phylogenies published)

VTR1B clearly forms a node with VTR1A. Additionally, in our 40-gene window analysis it is the second best syntenic hit for VTR1A in several comparisons, and in the species where VTR1B is not present or deleted (sea lamprey, gar), its synteny maps better to that of VTR1A.

Likewise, we aligned coelacanth VTR2A against all sea lamprey OTR-VTRs and compared introns and exons separately between VTR2A and the two sea lamprey OTR-VTRs with the highest max scores (OTR and VTR2C) (Figure S11). The alignment suggests that coelacanth VTR2A aligns better to sea lamprey OTR than to VTR2C, something that goes against the branching of VTR2A with the VTR2-subfamily in all the phylogenies. It does not agree either with our synteny results (40-100 gene window) where we found that VTR2A shows more synteny with the VTR2C-territory, in the species where VTR2A is deleted (e.g. coelacanth vs chicken). These findings further illustrate that sequence identity alone is not sufficient to infer the evolutionary history of some genes, where synteny clearly revealed them. Synteny combined with divergence rates provided stronger evidence for the designations we propose.

Assessing whole genome and segmental duplications to explain OTR-VTR evolution. Our SynMap2 dotplots (e.g. Figure S18) showed that the sea lamprey scaffold 10 shares significantly more synteny with chicken chromosome 1 and frog chromosome 3, while in human it shares more synteny with chromosome 7; as mentioned above, human chromosome 7 shows synteny with the territory of VTR2C found in non-mammalian species, and VTR1A with VTR2C are both located on chromosome 1 in chicken and chromosome 3 in frog (Figures 5 and 7; asterisks show statistically significant ($p<0.05$) difference between the first two chromosomes with the highest number of synteny hits). We found significant synteny between the sea lamprey scaffold 27 and human chromosome 3 (Figure 7), chicken chromosome 12, frog chromosome 4 and medaka chromosome 5, where OTR and VTR2B are located (Figure 6). There was not any significant mapping between the sea lamprey scaffold 49 with any chromosome in any of the species studied, but chromosome 6 in zebrafish, chromosome 4 in frog and chromosome 3 in human, where OTR-VTR2B are located, were the ones with the most hits (Figure 6, S5). This suggests that these two putative OTR-VTRs on scaffold 49 in lampreys were created by a segmental duplication of the OTR and VTR2B contig on scaffold 27. In support of this hypothesis, our within species SynMap2 dotplot of sea lamprey scaffold 49 against scaffolds 10 and 27 showed 11 and 52 hits respectively (Figure S18). In addition to that, the OTR-VTR genes on scaffold 49 show higher identity to OTR and VTR2B (Table S10; taking into account query cover which is very low for VTR1A and VTR2C).

The scaffolds where the OTR-VTRs are located according to sequence identity in hagfish did not show significant synteny with any scaffold/chromosome in any species, but this was partly expected because the reads are short and the evolutionary distance is longer (Figures S6-S9). Nevertheless, when these hagfish scaffolds were aligned against the genomes of other vertebrate species, they had the most gene hits in chromosomes where OTR and VTR2B or VTR1A and VTR2C were located in these species. Specifically, hagfish scaffold FYBX02010521.1 containing ENSEBUG00000001467 (its sequence clusters to VTR1A in the protein tree) showed most syntenic hits with the zebrafish and the human chromosomes where VTR1A is located, and the chicken chromosome where OTR is located (Figure S6); hagfish scaffold FYBX02010841.1 containing ENSEBUG00000007964 (its sequence clusters to VTR2C in the protein tree) showed synteny to only sea lamprey scaffold 10 (Figure S8), to chicken chromosome 1 where VTR2C is located (Figure S9), but also to the frog and human chromosomes where OTR is located (Figure S9).

Long non-coding RNA (lncRNA) synteny reveals divergences after duplications. We searched for other elements that may help identify syntenic regions in lampreys. We found one sequence of lncRNA next to each lamprey OTR, VTR1A and OT/VTR-49, and three next to VTR2C. We aligned them in all combinations within lamprey and found high identities between one of the three VTR2C lncRNAs (MSTRG.6000.1) when compared to OTR, VTR1A and

OT/VTR-49 lncRNAs. Another VTR2C-lncRNA also matched the OT/VTR-49-lncRNA. The lncRNAs next to OTR and VTR1A were also homologous. (Figure 9).

We searched for these lncRNAs in the human genome, but none of the lncRNAs flanking OTR-VTRs in sea lamprey display identity beyond threshold with any of the lncRNAs flanking human OTR-VTRs. This may not be surprising, since lncRNAs evolve rapidly, with >70% of lncRNAs having no sequence-similar orthologs in species separated by >50 million years of evolutionary divergence (Hezroni et al. 2015). Less than 100 lncRNAs have been traced to the last common ancestor of tetrapods and teleost fish (Hezroni et al. 2015). Almost all human lncRNAs we encountered showed high homology to each other, especially the ones next to VTR1A and VTR1B (Figure 10), but we were able to track them only to other mammalian species. An exception was AC078814.1 (located in between PPM1H and VTR1A) which was conserved in all species we studied down to the elephant shark, but not the sea lamprey. In fact, three genes that appear next to VTR1A across species (USP15-MON2-PPM1H) were not found in the synteny of sea lamprey VTR1A, nor on scaffold 10 where VTR1A is located. We found them on sea lamprey scaffold 22. We found the synteny of the surrounding genes unaltered in the human and zebrafish genomes and noted a gap where USP15- MON2-PPM1H genes were located in the lamprey (hg19 Chromosome 3: 50,002,983-50,087,955; zebrafish Chromosome 6: 53,206,414-53,263,985). This is evidence for a translocation event of these genes to VTR1A's territory post-lamprey.

A universal nomenclature for oxytocin and vasotocin receptor genes. Based on these findings, we propose a unified nomenclature for the OTR-VTRs. We propose that the naming follow the same root-names of the ligands, namely oxytocin (OTR) and vasotocin receptor (VTR) followed by enumerations (1A, 1B, 2A, 2B, 2C) that designate ancestral histories of the paralogous genes of the gene family. This would mean that all VTR1 receptors belong to the VTR1 family and all VTR2 receptors belong to the VTR2 family. The VTR1 family consists of VTR1A, VTR1B, and OTR; the VTR2 family consists of VTR2A, VTR2B, and VTR2C. We considered renaming the OTR to VTR1C, to be consistent with the standard. However, we thought this might be too radical a departure from the commonly named OTR. This is further justified in that although there is cross-talk in OT and VT binding to these receptors, OT is the dominant binding ligand for the OTR (Song and Albers 2018).

This universal vertebrate nomenclature would mean that what is commonly called arginine vasopressin receptor 1A (AVRP1A) in mammals, vasotocin receptor 4 (VT4) in birds, and vasotocin receptor (VasR) in frogs would all simply be called vasotocin receptor 1A (VTR1A; Table 2). What is commonly called oxytocin receptor (OXTR) in mammals, vasotocin receptor 3 (VT3) or mesotocin receptor (MTR) in birds and frogs, and isotocin receptor (ITR) in fish would all be called oxytocin receptor (OTR; Table 2). Similar diversity of previous and currently used names would change to a single name for the other four receptors, namely, VTR1B, VTR2A, VTR2B, and VTR2C (Table 2).

The evolutionary tree and synteny analyses indicate that OTR was the parent gene of VTR1A and VTR1B, and that VTR2B and VTR2C arose before VTR2A. In this manner, it would be appropriate to reorder the A, B, and C enumerations of these genes according to their predicted chronological order in evolution. However, we decided against this and to use previously used designations of them among the various terminologies, to have some consistency with some of the past literature and to not add to the confusion. Further, differences in the exon and protein coding trees in our study suggest that with more species, the exact chronological relationships could be revised, and therefore we erred on the side of caution.

Discussion

In this study, using synteny, sequence identity, gene family trees, and other analyses, we find strong evidence of orthology and paralogy of OT, VT, and six OTR and VTR receptors within vertebrates. For the ligands, we infer that vertebrate VT was inherited from a common ancestor with invertebrates, which then later gave rise to OT by way of a local translocated

duplication through DNA TE elements after the divergence of lampreys with other vertebrates. Further evidence that VT was the ancestral gene comes from several features OT displays, which are frequently encountered in locally duplicated genes, like intron shortening and increased intronic GC-content (Rayko, Jabbari, and Bernardi 2006). The VT and OT ligands were then maintained in all subsequent vertebrate divergences. We don't see surviving VT paralogs from the 1R WGD at the origin of vertebrates. We also identified the origin of the change in the orientation of OT, in the stem of all mammals, since it is already flipped in monotremes, earlier than what Gwee et al. 2009 had hypothesized. We also noted for the first time that spotted gar presents an OT inversion too; this inversion is species-specific, order-specific (Lepisosteiformes) or class-specific (holostei), since the teleost included in this study do not present such an inversion. Inversions are common to help scatter duplicated genes along a particular chromosomal arm or between chromosomes, with possible functional implications (Puerma, Orengo, and Aguadé 2016), but it is still unknown why such inversions takes place in a locally duplicated gene, long after it got duplicated (Furuta et al. 2011).

For the receptors, we infer that a VTR in invertebrates was passed onto a vertebrate ancestor, followed by a tandem duplication that gave rise to VTR1 and VTR2 paralogous genes. Thereafter, in a 1R WGD, one copy of the tandem gave rise to OTR and VTR2B, respectively, and the other copy gave rise to VTR1A and VTR2C, respectively (Figure 14, 1st and 2nd scenaria). In subsequent duplications, a VTR1 gave rise to VTR1B and a VTR2 gave rise to VTR2A. These subsequent duplications could have occurred by segmental duplications of ancestral chromosomes (Figure 14, 1st scenario) or a 2R of WGD (Figure 14, 2nd scenario). In either case, the orthology and evolution findings allowed us to propose a universal vertebrate-wide (and possibly invertebrate) nomenclature for the OT/VT and receptor families. This universal nomenclature will make it easier to translate findings and design experiments across vertebrates.

In our 1st proposed scenario (Figure 14), after the 1R of WGD in the vertebrate ancestor, a segmental duplication in the gnathostome ancestor gave rise to VTR1B from VTR1A, and second segmental duplication in the bony fish ancestor gave rise to VTR2A, most likely from VTR2C. Segmental duplications are not rare at these evolutionary time points, as other gene families have been proposed to have been expanded before the gnathostomes, such as the glycoprotein hormones and receptors (Buechi and Bridgham 2017), or before the bony fish, such as the secretory calcium-binding phosphoprotein (SCPP) gene family (Venkatesh et al. 2014). The later duplication of VTR2A maybe explains why it is the only gene in this family that induces cAMP signaling, when the rest of the OT-VTRs induce Ca2+ signaling (Birnbaumer 2002; Konno et al. 2010; Yamaguchi et al. 2012).

A 2nd scenario where after the 1R of WGD, a 2R occurred at some point in the evolution of jawed vertebrates is not excluded, but it is less supported by our findings. On the grounds that we encounter VTR1B in elephant shark, a 2R would have happened before the divergence of gnathostomes from cyclostomes (Figure 14, 2nd scenario). If a 2R had occurred at that point, then VTR2A would have been present already in the gnathostome ancestor before it was lost specifically in the elephant shark. We searched in the elephant shark for syntenies around a hypothetical deleted VTR2A, but we were not able to find any. Added to that, our intraspecies macrosynteny analysis revealed that of all the gene families neighboring human OTR-VTRs, only one family, SRGAP, presents all four gene-members. So according to a scenario invoking evolution by 2R of WGD followed by independent losses, we would have to assume many losses in the gene families surrounding OTR-VTRs, but also two losses in the OTR-VTR family itself: a VTR2 (i.e. a hypothetical VTR2D) located in the same chromosome as VTR1B, and an VTR1 (hypothetical VTR1C) in the same chromosome as VTR2A (Figure 4, 2nd scenario).

In contrast, the alternative hypothesis that the OTR-VTR gene family evolved through 1R of WGD plus segmental duplications requires fewer steps and fewer deletions (Figure 4, 1st scenario). According to Smith and Keinath 2015, a scenario of 1R of WGD plus segmental duplications is consistent with expectations given a simple random mutational model that

requires as few as six mutational steps (1R of WGD plus five segmental duplications/fissions), whereas models invoking 2R of WGD require between 12 and 18 steps. Smith et al. 2018 showed that the evolutionary history of the Hox gene family can be used as evidence for such a scenario (1R of WGD plus segmental duplications), and we propose that the OTR-VTR gene family brings further evidence as well. In this vein, our proposal for a specific gene family has greater repercussions on a wider and highly debated topic, that of the evolution of vertebrate genomes.

In any case, both our proposed scenaria differ from what has been proposed in previous studies (Lagman et al. 2013; Mayasich and Clarke 2016). Even though Mayasich and Clarke 2016 put forward a 1R of WGD plus segmental duplications scenario, they were not able to label each OTR-VTR they had identified in the sea lamprey somatic genome as specifically "OTR" or "VTR1A"; labels were left ambiguous (e.g. "OXTR/V1A/B"). This is because a 1:1 correspondence is very difficult to draw between orthologs at this level of evolution, due to whole-genome and whole-chromosome duplications (Smith et al. 2018). Despite these difficulties, we managed to obtain better resolution of orthology through our SynMap2 analysis at a chromosomal/superscaffold level and higher quality genome assemblies. According to this analysis, sea lamprey scaffolds 10 and 27 share significant synteny with other species' chromosomes where VTR1A with VTR2C and OTR with VTR2B reside, respectively. We did not find any significant mapping between the sea lamprey scaffold 49 and other species' chromosomes, meaning that the genes residing in this scaffold constitute species-specific duplications. Further, based on our synteny analysis in a 5-gene and a 40-100 gene windows, we designate PMZ_0003232-RA (on scaffold 27) as OTR, PMZ_0008155-RA (on scaffold 27) as VTR2B, PMZ_0013447-RA (on scaffold 10) as VTR1A and PMZ_0042163-RA (on scaffold 10) as VTR2C. The genes on scaffold 49 are most likely a segmental duplication from scaffold 27, as the greater synteny between scaffold 49 and scaffold 27, compared to scaffold 49 and scaffold 10 suggests (Figures S18).

According to the 2R of WGD scenario proposed by Lagman et al. 2013 and Mayasich and Clarke 2016, they occurred in the cyclostome ancestor. Mayasich and Clarke 2016 specifically note that VTR1B and VTR2A had evolved already in the cyclostome ancestor and were deleted in the sea lamprey. In our analysis of the germline sea lamprey genome, we did not find evidence for either VTR1B or VTR2A being lost. Furthermore, Smith et al. 2018 and Smith and Keinath 2015, as mentioned above, have proposed through different analyses that there was 1R of WGD that most likely occurred in the lamprey/cyclostome ancestor, not 2R. So according to our proposal, VTR1B and VTR2A came about post cyclostomes. Lastly, unlike Mayasich and Clarke 2016, we do not believe that a lamprey-specific 3R of WGD gave rise to the OTR-VTRs on scaffold 49 (they had identified only one of the two additional receptors), because of our synteny data that suggest these genes were a segmental duplication from scaffold 27.

Since our analyses include both hagfishes and lampreys, our findings can also help shed light on the specific timing of the 1R of WGD and, consequently, on the monophyly-paraphyly debate in cyclostomes (Smith et al. 2010). If 1R of WGD happened in the cyclostome ancestor in the stem of vertebrates, this means that lampreys and hagfishes belong to the same phylum, the cyclostomes, a monophyly scenario. If 1R of WGD occurred after divergence of hagfishes from lampreys, this means that lampreys and hagfishes represent two separate phyla, a paraphyly scenario. Smith et al. 2018 and Smith and Keinath 2015, who proposed the 1R of WGD-scenario, did not specify whether this 1R happened before or after the divergence with lampreys or in a cyclostome ancestor. Zhang et al. 2017 suggest that the 1R of WGD event happened after lampreys diverged from the ancestor with hagfishes, based on their findings of a single ParaHox gene cluster in two hagfish species (inshore hagfish (*Eptatretus burgeri*) and the Atlantic hagfish (*Myxine glutinosa*)) and two gene clusters in lampreys (Zhang et al. 2017).

Our findings in hagfish leave open some possibilities due to the assembly quality. We found two OTR-VTRs in inshore hagfish and four in sea lamprey (without taking into account what we think of as lamprey-specific segmental duplications). Since the hagfish-scaffolds are

relatively short (scaffold N50=2,692,996) and these two genes do not appear on the same scaffold, we cannot decipher if they correspond to two genes that are located in the same superscaffold in sea lamprey (like OTR and VTR2B, or VTR1A and VTR2C).

Our hypothesis is that the two OTR-VTRs we encounter in inshore hagfish were located in the same chromosome in the vertebrate ancestor (Figure 14), something that might still be the case in the genome of the inshore hagfish, once a chromosomal-assembly comes to light. We do not exclude a scenario in which these two OTR-VTRs were located in the same chromosome in the vertebrate ancestor which was fissioned in inshore hagfish. We cannot be sure whether these two receptors in hagfish were orthologous to either the OTR and VTR2B pair or the VTR1A and VTR2C pair, but our sequence identity and synteny analyses support the latter. We consider less likely the following scenaria: first, that there are four OTR-VTRs in the inshore hagfish genome in total, but two of them were not properly sequenced or not assembled due to scaffolding issues, and second, that two of them were lost specifically in hagfish. As far as we are aware, our study is the first time that the hagfish genome was used to attempt to decipher the evolutionary history of OTR-VTRs. We consider its inclusion crucial for understanding the origin and evolution of OTR-VTRs, since it represents the most ancient branch with other vertebrates.

The inclusion of 33 vertebrate genomes in our study was also crucial for our understanding of the chromosomal fusions and fissions in vertebrate evolution (for the chromosomes where OTR-VTRs are located). VTR1A and VTR2C appear on the same chromosome/scaffold in all vertebrate species except teleost fish and mammals. This concurs with reconstructions of putative ancestral tetrapod chromosomes (Uno et al. 2012) and of putative pre-teleost duplicated-chromosomes (Nakatani and McLysaght 2017), where the chromosomes where VTR1A and VTR2C are located figure in both cases as a single putative ancestral chromosome. This means that in the bony fish-ancestor these genes were located on the same chromosome, which was subjected to fissions independently in teleost fish and in mammals. Considering this, we can hypothesize and expect that when the coelacanth genome will be assembled at a chromosome-level, the scaffolds where we find VTR1A and VTR2C will belong to the same chromosome.

The presence of both OT and VT in all jawed vertebrates with OT commonly proposed to function mainly in social behaviors and reproduction, and VT in osmoregulation, could suggest that these functions are hallmarks of vertebrates. However, the two peptides have been shown to have a much wider variety of overlapping as well as other distinct functions (Caldwell and Young 2006), making it difficult to hypothesize a specific function that led to strong selection for maintenance of both genes in vertebrates. Interestingly, although OT is not present in lampreys, they do have an OTR (and another OTR-like duplication). It is possible that VT in lampreys acts through these OTRs, as it happens in elephant shark where both OT and VT were shown to activate OTR to a similar extent; a greater response of OTR to OT than VT is found for the first time in teleost fish (Yamaguchi et al. 2012). This is logical, given the sister relationships of the ligands and receptors.

For the receptors, the finding that all jawed vertebrates since the divergence with lampreys have 4 to 5 of the 6 OTR-VTR receptors, in different combinations, indicates that having all 6 may be a disadvantage. All vertebrates have the VTR1 receptors OTR and VTR1A. This suggests that once OTR evolved at the origin of vertebrates, the OT ligand post-lamprey divergence may have become dependent on the presence of OTR, and VT on VTR1A. Of the remaining receptors, 1 to 2 of them have been lost in specific vertebrate lineages, indicating that they are less essential for vertebrate survival.

Conclusions

In this study we put forward a mechanism via which the OT and VT ligands evolved from each other by local duplication, after the divergence of gnathostomes from cyclostomes, i.e. via DNA transposable elements. Further, we propose an evolutionary scenario according to

which the OTR-VTRs emerged through 1R of WGD followed by segmental duplications, although we don't fully exclude a 2R scenario. Based on these findings and synteny analyses, we propose a new unified nomenclature for the OT-VT ligand and receptor gene families for all vertebrates.

Methods

Synteny analyses. In order to define orthology in the OT-VT and OTR-VTRs in all vertebrates (and crucially in the sea lamprey), we employed inter-species synteny analyses at three different scales: a 10-gene window (via BLAT/BLAST searches, SynFind and GeVo); a 40-100-gene window (via SynFind); and at the chromosomal level with syntenic dotplots (via SynMap2). To further trace their evolutionary history, we employed intra-genome macrosynteny analysis for the human genome (intraspecies), searching for paralogous genes at a 10 Mb window around human OTR-VTRs. Microsynteny was useful for determining orthologous and paralogous relationships between genes in the majority of the vertebrate lineages. Macrosynteny was useful for determining orthologous and paralogous relationships between genes found in lampreys and hagfish and the rest of the vertebrates, for uncovering the evolutionary history of the receptors and for confirming our microsynteny findings in a larger window. Below we describe the specific methods for each approach.

Microsynteny analysis between species (10-gene window). We ran microsynteny analysis around the OT, VT, VTR1A, VTR1B, VTR2A, VTR2B and VTR2C regions, manually scanning annotations for 5 protein coding genes before and after each gene (Table S1) in 33 species spanning all major vertebrate lineages (Table 1). The candidate genes in each species, whether they were annotated as OTR-VTRs or not, were first selected by BLAT and BLAST searches using the UCSC genome browser (http://genome.ucsc.edu/) (Kent 2002) and the SynFind tool from the CoGe comparative genomics research platform (Lyons and Freeling 2008). The NCBI RefSeq database and Ensembl (Zerbino et al. 2018) prediction programs were used to identify the neighboring genes (the NCBI accession number and the Ensemble or Gene ID for each gene in each species can be found in Table S1). We used the aliases in NCBI and Ensembl for each gene in each organism and listed the most frequent ones in Table 2. For the Japanese lamprey (*Lethenteron japonicum*) we used the synteny data available in (Mayasich and Clarke 2016); for the sea lamprey (*Petromyzon marinus*) we used the assembly of the germline genome (Smith et al. 2018), analyzing it with BLAST, Genome Browser and Gene Search tools (https://genomes.stowers.org/organism/Petromyzon/marinus). When our target genes appeared to be lost in the species' genome (no initial BLAST hit), we searched the surrounding gene territory to determine whether only the receptor of interest or a larger block of genes was deleted, or whether the deletion appeared to be due to an incomplete genome assembly or assembly artifact For teleost-specific duplications, we ran microsynteny analysis in the zebrafish (*Danio rerio*) as a teleost-representative (Table S3).

We also used the inshore hagfish genome (*Eptatretus burgeri*), but since the contigs were short and not fully annotated, we first BLAT searched in Ensembl using all OT-VT and OTR-VTR sequences of all the aforementioned species against the hagfish genome; we found two putative OTR-VTRs in two separate contigs in the hagfish assembly and we used the 'Region comparison' tool of Ensembl to map each gene of these contigs against the human, zebrafish and lamprey genomes (Table S5). BLAST did not bring any results for OT-VT ligands, so we used the 'Gene Tree' tool that constructs a phylogeny with all the orthologous and paralogous genes of a gene-family, using the sea lamprey VT as reference (http://www.ensembl.org/Multi/GeneTree/Image?collapse=none;db=core;gt=ENSGT00390000004511).

Macrosynteny analysis between species (40-100 gene window). Dotplots of sequence alignments between pairs of species were generated using SynFind (Lyons and Freeling 2008). Two organisms were compared with the default Last alignment parameters. This produced a data matrix with each gene from the query organism alongside all matching genes in the reference organism. This data matrix was parsed and analyzed using a custom R script. Then a 40-gene window centered around a given gene (i.e. receptor) in the reference organism was searched and identified (x axis; Figure 4, Figure S3). ln most comparisons the syntenic blocks leveled out within the 40 gene window, but for the species representing older divergences, sea lamprey and coelacanth in particular, our cumulative lines were still going up at the 40-gene limit, so we extended these windows to see how large the syntenic blocks were. As we move 5' (left) or 3' (right) from zero (the focus gene), the value of each line will increase with number of matching genes if there are additional genes in synteny, visualizing stretches of synteny on either side of the focus gene. This allows us to see large stretches of homologous sequences that may be interspersed by divergent sequences. We ran this data matrix test using the OTR-VTRs from at least one species per major vertebrate class, namely the sea lamprey (*Petromyzon marinus*), coelacanth (*Latimeria chalumnae*), elephant shark (*Callorhinchus milii*), spotted gar (*Lepisosteus oculatus*), western clawed frog (*Xenopus tropicalis*), chicken (*Gallus gallus*), and human (*Homo sapiens*), in all possible combinations. We also BLAST searched the OTR-VTRs of sea lamprey, chicken and human against the OTR-VTR of the other species (e.g. coelacanth, elephant shark, spotted gar, frog, chicken and human), in order to understand relationships between the percent identity/divergence and synteny (Table S6). For this last BLAST analysis, only results with a bitscore >40 and hits with high probability E-value < 10^{-4} were kept.

Macrosynteny analysis between species (Chromosome/contig-window). We used SynMap2 (Haug-Baltzell et al. 2017) to generate syntenic dotplots of sequence alignments between the sea lamprey andthe inshore hagfish contigs/scaffolds that contain OTR-VTRs against the Japanese medaka (*Oryzias latipes*), zebrafish, frog, chicken and human genomes (Figure S5 for an example between chicken and lamprey). SynMap2 identifies collinear sets of putative genes or regions of sequence similarity to infer synteny between two sequences and generates a dotplot of the results. We used the default parameters (as of December 2018), except for 'Minimum number of aligned pairs'. This parameter defines the minimum number of homologous genes (based on Last default parameters) that should be found at a 20-gene distance from each other, for these genes to be considered syntenic and to appear on the dotplot. We selected 3 as a minimum number when we compared the sea lamprey against the Japanese medaka, zebrafish, frog and chicken genomes; both 2 and 3 for the human genome; 2 when we mapped the hagfish contigs against the sea lamprey, Japanese medaka, zebrafish, frog, chicken and human genomes, since the hagfish contigs were very short. For the hagfish, we also ran a dotplot with 1 as minimum number to search for all possible homologous hits, regardless of synteny.

To test for significant differences, we ran a t-test on analyses between the first two chromosomes with the highest number of hits, using the number of genes in each sea lamprey super-scaffold studied to calculate mean values. For the cases that reached significance, in order to confirm that the number of hits we found was independent from the number of protein-coding genes located on each chromosome, we applied a gene density-normalisation test, by dividing the number of hits between each sea lamprey super-scaffold by the number of protein coding genes on each of the Japanese medaka, zebrafish, frog, chicken or human chromosomes with most hits.

Macrosynteny analysis within species (10 Mb window). We ran an intra-species macrosynteny analysis in the human genome in order to assess the gene-families that appear in

synteny in the neighboring regions of human OTR-VTRs and in the territories from where OTR-VTRs were deleted in the human genome. We used human, as it was the best assembled genome and therefore subject to generating less errors and still representative of vertebrates generally, and mammals specifically. We listed all genes found in a 10 Mb window from the OTR-VTRs that are present in the human genome (OTR, VTR1A, VTR1B, VTR2A). We chose a 10 Mb window, as we noted that this genomic region size captured in many cases macrosynteny of > 40 genes between species in our macrosynteny analyses described above, allowing the two analysis to be comparable. We then searched each gene in the HGNC Database (https://www.genenames.org/) in order to classify it in its gene-family according to the HUGO Gene Nomenclature Committee. We performed the same kind of analysis for the territories from which OTR-VTRs were found in non-human species (VTR2B, VTR2C) that have appeared to have been deleted from the human (and other mammalian) genome (Tables S7-S8). We defined these territories by manually identifying in the human genome the genes around spotter gar VTR2B and chicken VTR2C (Figures S14-17); some of these syntenies around the deleted OTR-VTRs were previously identified in (Lagman et al. 2013 and Yamaguchi et al. 2012). For each gene family identified in the 10 Mb window around each OTR-VTR focus gene, we searched the remaining gene-members in the genome and included them in our list (Tables S7-S8). In a few rare cases, we found orthologous/paralogous gene family members outside of the 10 Mb window, but on the same chromosome (non-shaded genes in Table S8).

Evolutionary history analyses of the OT-VT genomic region. We had noted annotated DNA transposable elements in the UCSC annotated genome in close vicinity of the OT-VT genes. Thus, we quantitatively searched for DNA transposable elements (TEs) around the OT and VT region in the human (Figure S1) and chimpanzee genomes using the RepeatMasker tool in the UCSC Genome Browser (see Suppl. Material) and we obtained information for each specific TE via Dfam 2.0 (Hubley et al. 2016). We calculated the GC content using http://www.endmemo.com/bio/gc.php. We aligned the introns of human OT and VT in all possible combinations using DIALIGN (Morgenstern et al. 2006) and compared the length of the introns with the higher identity (first intron of OT vs. first intron of VT) using the Serial Cloner. The elephant shark genome was not annotated for DNA transposable elements, so we were not able to trace the TEs in this species. We only compared intron length and GC content between OT and VT. The information on the orientation of OT and VT in all the species studied is available in Table S2. In this report, we added the orientation of also the Kangaroo rat (*Dipodomys ordii;* Dord_2.0) OT-VT.

Evolutionary history analyses of the OTR-VTRs. In order to assess how the OTR-VTR family initially emerged and, more specifically, whether the OTR-VTRs originate back to the same ancestral chromosome, we followed the common practice of mapping our regions of interest back to putative ancestral chromosomes, reconstructed via chromosome fragments with reliable synteny. We followed four different ancestral models, whose synteny was based on different species and different genome-qualities (Table S9) (Nakatani et al. 2007; Putnam et al. 2008; Smith et al. 2013, 2018).

We searched for the presence of annotated OTR-VTRs in four outgroup invertebrate lineages (literature review and BLAST/BLAT searches), namely in sea squirt (*Ciona intestinalis*), roundworm (*Caenorhabditis elegans*), pond snail (*Lymnaea stagnalis*), and amphioxus (*Branchiostoma floridae*). For the amphioxus genome (B. floridae v2.0) we performed BLAT queries with the OTR-VTR FASTA sequences from all the species studied using the JGI genome browser: https://genome.jgi.doe.gov/pages/blast-query.jsf?db=Brafl1.

In order to test which sea lamprey scaffold most likely carried the orthologous ancestral gene(s) that was predicted to be duplicated in the 1R of WGD, we compared via BLASTn (same parameters) the sea lamprey OTR-VTRs in all possible combinations. Thereafter, we compared

the exons and introns of the identified genes separately to understand the divergence of the duplicate genes in exon-intron structure, following the paradigm proposed in (Xu et al. 2012). The max score and percent identities of the comparisons that were above threshold are shown in Figures S12-S13. We then performed a similar analysis for the VTR1B and VTR2A in elephant shark andcoelacanth respectively against sea lamprey and human to test if sequence identity can help solve ancestry questions in these duplicated genes. Those that yielded the first two highest identities are shown in Figures S10-S11. In order to shed light on the orthology between the inshore hagfish and the sea lamprey OTR-VTRs, we compared their exons and introns separately as well.

Finally, we analyize conserved non-coding RNA synteny around the OTR-VTRs, we looked for them in alignments in all the species studied in Ensembl, in the miRbase (http://www.mirbase.org/; miRbase 22 release)), and the miRviewer (Kiezun et al. 2012; last update of the database: Feb 28, 2012). We aligned (BLASTn) long non-coding RNA regions within species in sea lamprey and human (Figures 9 and 10).

Gene tree phylogeny analyses.

Exonic-tree. Exonic sequences from all the OT-VTRs in sea lamprey, elephant shark, coelacanth, spotted gar, zebrafish, anole lizard, alligator, frog, turtle, chicken, zebra finch, mouse and human for vertebrates and amphioxus as an invertebrate outgroup were aligned with MAFFT. We used settings appropriate for high divergence sequences with the potential for large gaps surrounding regions of conservation. Any OTR-VTR less than 100 bp was excluded, as alignments on such short sequences are unreliable, giving weakly supported gene trees. From this alignment, we generated a Phylogenetic Maximum Likelihood tree using RAXML with 100 bootstrap replicates (Figure 11).

Protein-tree. A second maximum likelihood phylogenetic tree was constructed with the 'Gene tree' tool in Ensembl (Gene Tree ID: ENSGT00760000119156): gene trees are constructed using one representative protein sequence for every gene in every species in Ensembl. The trees were generated by the Gene Orthology/Paralogy prediction method pipeline (https://www.ensembl.org/info/genome/compara/homology_method.html). We manually curated the Ensembl tree using the unified nomenclature we proposed (Figure 12). A version of the fully expanded tree can be accessed here: http://jul2018.archive.ensembl.org/Multi/GeneTree/Image?collapse=5527114%2C5527064%2C5526527%2C5526321%2C5526733%2C5526537%2C5526761%2C5526759%2C5526979%2C5526590%2C5526778%2C5526518%2C5526927%2C5526552%2C5526279%2C5527074%2C5527228%2C5526122%2C5526479%2C5526205%2C5527217%2C5526920%2C5526768;db=core;gt=ENSGT00760000119156;gtr=class

References

Acher, Roger, and Jacqueline Chauvet. 1988. "Structure, Processing and Evolution of the Neurohypophysial Hormone-Neurophysin Precursors." *Biochimie* 70(9): 1197–1207. http://linkinghub.elsevier.com/retrieve/pii/030090848890185X.

Baeyens, Dennis A., and Lawrence E. Cornett. 2006. "The Cloned Avian Neurohypophysial Hormone Receptors." *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 143(1): 12–19. http://linkinghub.elsevier.com/retrieve/pii/S1096495905002174.

Bailey, Jeffrey A., Ge Liu, and Evan E. Eichler. 2003. "An Alu Transposition Model for the Origin and Expansion of Human Segmental Duplications." *The American Journal of Human Genetics* 73(4): 823–34. http://linkinghub.elsevier.com/retrieve/pii/S0002929707636313.

Birnbaumer, Mariel. 2002. “Vasopressin Receptors.” In *Hormones, Brain and Behavior*, Elsevier, 803–10. http://linkinghub.elsevier.com/retrieve/pii/B9780125321044500603.

Buechi, Hanna B., and Jamie T. Bridgham. 2017. “Evolution of Specificity in Cartilaginous Fish Glycoprotein Hormones and Receptors.” *General and Comparative Endocrinology* 246: 309–20. https://linkinghub.elsevier.com/retrieve/pii/S0016648017300072.

Caldwell, H. K., & Young, W. S. 2006. “Oxytocin and Vasopressin: Genetics and Behavioral Implications.” In *Handbook of Neurochemistry and Molecular Neurobiology*, Springer US, 573–607.

Furuta, Y. et al. 2011. “Birth and Death of Genes Linked to Chromosomal Inversion.” *Proceedings of the National Academy of Sciences* 108(4): 1501–6. http://www.pnas.org/cgi/doi/10.1073/pnas.1012579108.

Ghiurcuta, C. G., and B. M. E. Moret. 2014. “Evaluating Synteny for Improved Comparative Studies.” *Bioinformatics* 30(12): i9–18. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu259.

Gubrij, Konstantin I. et al. 2005. “Molecular Cloning of an Oxytocin-like Receptor Expressed in the Chicken Shell Gland.” *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 142(1): 37–45. http://linkinghub.elsevier.com/retrieve/pii/S1096495905001120.

Gwee, Pai Chung, Boon Hui Tay, Sydney Brenner, and Byrappa Venkatesh. 2009. “Characterization of the Neurohypophysial Hormone Gene Loci in Elephant Shark and the Japanese Lamprey: Origin of the Vertebrate Neurohypophysial Hormone Genes.” *BMC Evolutionary Biology* 9(1): 1–15.

Haug-Baltzell, Asher et al. 2017. “SynMap2 and SynMap3D: Web-Based Whole-Genome Synteny Browsers.” *Bioinformatics* 33(14): 2197–98.

Hezroni, Hadas et al. 2015. “Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species.” *Cell Reports* 11(7): 1110–22. https://linkinghub.elsevier.com/retrieve/pii/S1611124715004106.

Hoyle, Charles HV. 1999. “Neuropeptide Families and Their Receptors: Evolutionary Perspectives.” *Brain research* 848(1–2): 1–25.

Hubley, Robert et al. 2016. “The Dfam Database of Repetitive DNA Families.” *Nucleic Acids Research* 44(D1): D81–89. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1272.

Jaillon, Olivier et al. 2004. “Genome Duplication in the Teleost Fish Tetraodon Nigroviridis Reveals the Early Vertebrate Proto-Karyotype.” *Nature* 431(7011): 946–57. http://www.nature.com/doifinder/10.1038/nature03025.

Jarvis, E. D. et al. 2014. “Whole-Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds.” *Science* 346(6215): 1320–31. http://www.sciencemag.org/cgi/doi/10.1126/science.1253451.

Kent, W. J. 2002. “BLAT---The BLAST-Like Alignment Tool.” *Genome Research* 12(4): 656–64. http://www.genome.org/cgi/doi/10.1101/gr.229202.

Kiezun, Adam et al. 2012. “MiRviewer: A Multispecies MicroRNA Homologous Viewer.” *BMC Research Notes* 5(1): 92. http://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-5-92.

Knobloch, H. Sophie, and Valery Grinevich. 2014. “Evolution of Oxytocin Pathways in the Brain of Vertebrates.” *Frontiers in Behavioral Neuroscience* 8(February): 1–13.

http://journal.frontiersin.org/article/10.3389/fnbeh.2014.00031/abstract.

Konno, Norifumi et al. 2010. “Molecular Cloning and Characterization of V2-Type Receptor in Two Ray-Finned Fish, Gray Bichir, Polypterus Senegalus and Medaka, Oryzias Latipes.” *Peptides* 31(7): 1273–79. http://linkinghub.elsevier.com/retrieve/pii/S0196978110001683.

Korlach, Jonas et al. 2017. “De Novo PacBio Long-Read and Phased Avian Genome Assemblies Correct and Add to Reference Genes Generated with Intermediate and Short Reads.” *GigaScience* 6(10). https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/gix085/4096262.

Lagman, David et al. 2013. “The Vertebrate Ancestral Repertoire of Visual Opsins, Transducin Alpha Subunits and Oxytocin/Vasopressin Receptors Was Established by Duplication of Their Shared Genomic Region in the Two Rounds of Early Vertebrate Genome Duplications.” *BMC Evolutionary Biology* 13(1).

Leung, Cary H. et al. 2011. “Neural Distribution of Vasotocin Receptor MRNA in Two Species of Songbird.” *Endocrinology* 152(12): 4865–81. https://academic.oup.com/endo/article-lookup/doi/10.1210/en.2011-1394.

Liutkeviciute, Zita et al. 2016. “Global Map of Oxytocin/Vasopressin-like Neuropeptide Signalling in Insects.” *Scientific Reports* 6(1): 39177. http://www.nature.com/articles/srep39177.

Lovell, Peter V et al. 2014. “Conserved Syntenic Clusters of Protein Coding Genes Are Missing in Birds.” *Genome Biology* 15(12): 565. http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0565-1.

Lyons, Eric, and Michael Freeling. 2008. “How to Usefully Compare Homologous Plant Genes and Chromosomes as DNA Sequences.” *The Plant Journal* 53(4): 661–73. http://doi.wiley.com/10.1111/j.1365-313X.2007.03326.x.

Mayasich, Sally A., and Benjamin L. Clarke. 2016. “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).” *General and Comparative Endocrinology* 226: 88–101. http://linkinghub.elsevier.com/retrieve/pii/S0016648016300028.

Meyer-Lindenberg, Andreas, Gregor Domes, Peter Kirsch, and Markus Heinrichs. 2011. “Oxytocin and Vasopressin in the Human Brain: Social Neuropeptides for Translational Medicine.” *Nature Reviews Neuroscience* 12(9): 524–38. http://dx.doi.org/10.1038/nrn3044.

Morgenstern, B., Prohaska S.J., Pöhler, D., Stadler, P.F. 2006. “Multiple Sequence Alignment with User-Defined Anchor Points.” *Algorithms for Molecular Biology* 1(6).

Nakatani, Yoichiro, and Aoife McLysaght. 2017. “Genomes as Documents of Evolutionary History: A Probabilistic Macrosynteny Model for the Reconstruction of Ancestral Genomes.” *Bioinformatics* 33(14): i369–78. https://academic.oup.com/bioinformatics/article/33/14/i369/3953974.

Nakatani, Yoichiro, Hiroyuki Takeda, Yuji Kohara, and Shinichi Morishita. 2007. “Reconstruction of the Vertebrate Ancestral Genome Reveals Dynamic Genome Reorganization in Early Vertebrates.” *Genome Research* 17(9): 1254–65.

Ocampo Daza, Daniel, Michalina Lewicka, and Dan Larhammar. 2012. “The Oxytocin/Vasopressin Receptor Family Has at Least Five Members in the Gnathostome Lineage, Inclucing Two Distinct V2 Subtypes.” *General and Comparative Endocrinology* 175(1): 135–43. https://linkinghub.elsevier.com/retrieve/pii/S0016648011003960.

Ohno, Susumu. 1970. “Evolution by Gene Duplication.” In New York: Springer-Verlag.

Ondrasek, Naomi R., Sara M. Freeman, Karen L. Bales, and Rebecca M. Calisi. 2018. "Nonapeptide Receptor Distributions in Promising Avian Models for the Neuroecology of Flocking." *Frontiers in Neuroscience* 12. https://www.frontiersin.org/article/10.3389/fnins.2018.00713/full.

Parry, Laura J., Ross A.D. Bathgate, and Richard Ivell. 2000. "Mammalian Mesotocin: CDNA Sequence and Expression of an Oxytocin-like Gene in a Macropodid Marsupial, the Tammar Wallaby." *General and Comparative Endocrinology* 118(2): 187–99. http://linkinghub.elsevier.com/retrieve/pii/S0016648000974641.

Puerma, Eva, Dorcas J. Orengo, and Montserrat Aguadé. 2016. "The Origin of Chromosomal Inversions as a Source of Segmental Duplications in the Sophophora Subgenus of Drosophila." *Scientific Reports* 6(1): 30715. http://www.nature.com/articles/srep30715.

Putnam, Nicholas H. et al. 2008. "The Amphioxus Genome and the Evolution of the Chordate Karyotype." *Nature* 453(7198): 1064–71.

Rayko, Edda, Kamel Jabbari, and Giorgio Bernardi. 2006. "The Evolution of Introns in Human Duplicated Genes." *Gene* 365: 41–47. http://linkinghub.elsevier.com/retrieve/pii/S0378111905006475.

Smith, J. J., N. R. Saha, and C. T. Amemiya. 2010. "Genome Biology of the Cyclostomes and Insights into the Evolutionary Biology of Vertebrate Genomes." *Integrative and Comparative Biology* 50(1): 130–37. https://academic.oup.com/icb/article-lookup/doi/10.1093/icb/icq023.

Smith, Jeramiah J. et al. 2013. "Sequencing of the Sea Lamprey (Petromyzon Marinus) Genome Provides Insights into Vertebrate Evolution." *Nature Genetics* 45(4): 415–21. http://dx.doi.org/10.1038/ng.2568.

———. 2018. "The Sea Lamprey Germline Genome Provides Insights into Programmed Genome Rearrangement and Vertebrate Evolution." *Nature Genetics* 50(2): 270–77. http://dx.doi.org/10.1038/s41588-017-0036-1.

Smith, Jeramiah J., and Melissa C. Keinath. 2015. "The Sea Lamprey Meiotic Map Improves Resolution of Ancient Vertebrate Genome Duplications." *Genome Research* 25(8): 1081–90.

Song, Zhimin, and H. Elliott Albers. 2018. "Cross-Talk among Oxytocin and Arginine-Vasopressin Receptors: Relevance for Basic and Clinical Studies of the Brain and Periphery." *Frontiers in Neuroendocrinology* 51: 14–24. https://linkinghub.elsevier.com/retrieve/pii/S009130221730064X.

Theofanopoulou, Constantina, Cedric Boeckx, and Erich D. Jarvis. 2017. "A Hypothesis on a Role of Oxytocin in the Social Mechanisms of Speech and Vocal Learning." *Proceedings of the Royal Society B: Biological Sciences* 284(1861): 20170988. http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2017.0988.

Uno, Yoshinobu et al. 2012. "Inference of the Protokaryotypes of Amniotes and Tetrapods and the Evolutionary Processes of Microchromosomes from Comparative Gene Mapping" ed. Dirk Steinke. *PLoS ONE* 7(12): e53027. https://dx.plos.org/10.1371/journal.pone.0053027.

Venkatesh, Byrappa et al. 2014. "Elephant Shark Genome Provides Unique Insights into Gnathostome Evolution." *Nature* 505(7482): 174–79. http://www.nature.com/articles/nature12826.

Wicker, Thomas et al. 2007. "A Unified Classification System for Eukaryotic Transposable Elements." *Nature reviews. Genetics* 8(12): 973–82. http://www.ncbi.nlm.nih.gov/pubmed/17984973.

Xu, G., C. Guo, H. Shan, and H. Kong. 2012. "Divergence of Duplicate Genes in Exon-Intron Structure." *Proceedings of the National Academy of Sciences* 109(4): 1187–92.

http://www.pnas.org/cgi/doi/10.1073/pnas.1109047109.

Yamaguchi, Yoko et al. 2012. “The Fifth Neurohypophysial Hormone Receptor Is Structurally Related to the V2-Type Receptor but Functionally Similar to V1-Type Receptors.” *General and Comparative Endocrinology* 178(3): 519–28. http://dx.doi.org/10.1016/j.ygcen.2012.07.008.

Yamashita, Kaoru, and Takashi Kitano. 2013. “Molecular Evolution of the Oxytocin–oxytocin Receptor System in Eutherians.” *Molecular Phylogenetics and Evolution* 67(2): 520–28. https://linkinghub.elsevier.com/retrieve/pii/S1055790313000742.

Yun, Seongsik et al. 2015. “Prevertebrate Local Gene Duplication Facilitated Expansion of the Neuropeptide GPCR Superfamily.” *Molecular Biology and Evolution* 32(11): 2803–17.

Zerbino, Daniel R et al. 2018. “Ensembl 2018.” *Nucleic Acids Research* 46(D1): D754–61. http://academic.oup.com/nar/article/46/D1/D754/4634002.

Zhang, Huixian et al. 2017. “Lampreys, the Jawless Vertebrates, Contain Only Two ParaHox Gene Clusters.” *Proceedings of the National Academy of Sciences* 114(34): 9146–51. http://www.pnas.org/lookup/doi/10.1073/pnas.1704457114.

Tables

		Organism	Assemblies used in this study	GenBank assembly accession
Vertebrates	**Mammals**	Human (*Homo Sapiens*)	GRCh38.p12	GCA_000001405.27
		Chimpanzee (*Pan troglodytes*)	Clint_PTRv2 and panTro3.0	GCA_002880755.3 and GCA_000001515.5
		Western Gorilla (*Gorilla gorilla*)	gorGor5	GCA_000151905.3
		Northern white-cheeked Gibbon (*Nomascus leucogenys*)	Nleu_3.0	GCA_000146795.3
		Rhesus Macaque (*Macaca mulatta*)	Mmul_8.0.1	GCA_000772875.3
		Marmoset (*Callithrix jacchus*)	Callithrix_jacchus-3.2 and ASM275486v1	GCA_000004665.1 and GCA_000832365.1
		Mouse lemur (*Microcebus murinus*)	Mmur_3.0	GCA_000165445.3
		Mouse (*Mus musculus*)	GRCm38.p6	GCA_000001635.8
		Prarie Vole (*Microtus ochrogaster*)	MicOch1.0	GCA_000317375.1
		Cow (*Bos taurus*)	ARS-UCD1.2_and UMD3.1.1	GCA_002263795.2 and GCA_000003055.5
		Yangtze River Dolphin (*Lipotes vexillifer*)	Lipotes_vexillifer_v1	GCA_000442215.1
		Horse (*Equus caballus*)	EquCab3.0	GCA_002863925.1
		Dog (*Canis lupus familiaris*)	CanFam3.1	GCA_000002285.2
		Megabat (*Pteropus vampyrus*)	pteVam1 and Pvam_2.0	GCA_000151845.1 and GCA_000151845.2
		Platypus (*Ornithorhynchus anatinus*)	OANA5	GCF_000002275.
	Birds	Chicken (*Gallus gallus*)	GRCg6a	GCA_000002315.5
		Anna's hummingbird (*Calypte anna*)	ASM69908v1	GCA_000699085.1
		Zebra finch (*Taeniopygia guttata*)	taeGut3.2.4 and MUGN00000000	GCA_000151805.2 and GCA_002008985.2
	Reptiles	American Alligator (*Alligator mississippiensis*)	ASM28112v4	GCA_000281125.4
		Carolina anole-lizard (*Anolis carolinensis*)	AnoCar2.0	GCA_000090745.2
		Painted turtle (*Chrysemys picta bellii*)	Chrysemys_picta_bellii-3.0.3	GCA_000241765.2
	Amphibia	Tropical clawed frog (*Xenopus tropicalis*)	JGI 4.2 and Xenopus_tropicalis_v9.1	GCA_000004195.1 and GCA_000004195.3
	Holost fish	Spotted Gar (*Lepisosteus oculatus*)	LepOcu1	GCA_000242695.1
	Teleost fish	Southern platyfish (*Xiphophorus maculatus*)	X_maculatus-5.0-male	GCA_002775205.2
		Three-spined stickleback (*Gasterosteus aculeatus*)	BROAD S1	GCA_000180675.1
		Japanese medaka (*Oryzias latipes*)	ASM223467v1	GCA_002234675.1
		Nile Tilapia (*Oreochromis niloticus*)	O_niloticus_UMD_NMBU and Orenil1.0	GCA_001858045.3 and GCA_000188235.2
		Zebrafish (*Danio rerio*)	GRCz11	GCA_000002035.4
	Coelacanths	Coelacanth (*Latimeria chalumnae*)	LatCha1	GCA_000225785.1
	Sharks	Elephant Shark (*Callorhinchus milii*)	Callorhinchus_milii-6.1.3	GCA_000165045.2
	Lampreys	Japanese lamprey (*Lethenteron japonicum*)	LetJap1.0	GCA_000466285.1
		Sea Lamprey (*Petromyzon marinus*)	gPmar1.0	GCA_002833325.1
	Hagfishes	Inshore hagfish (*Eptatretus burgeri*)	Eburgeri_3.2	GCA_900186335.2
Invertebrates	**Tunicates**	Sea squirt (*Ciona intestinalis*)	KH	GCA_000224145.2
	Lancelets	Amphioxus (*Branchiostoma floridae*)	B. floridae v2.0	GCA_000003815.1
	Mollusks	Great pond snail (*Lymnaea stagnalis*)	v1.0	GCA_900036025.1
	Nematodes	Roundworm (*Caenorhabditis elegans*)	Cael_CB4856_1.0	GCA_000975215.1

Table 1: Species list of genome assemblies used in this study. Specific assembly file names and GCA accession numbers are listed.

Mammals	Birds	Turtles/ Crocodiles	Frogs	Fish	Elephant Shark	Universal Vertebrate Revision
Oxytocin (OXT, OT, Oxy) Neurophysin (NPI) Mesotocin (MT)	Mesotocin (MT, MST), Oxt-like, Neurophysin-1-like	Mesotocin (MT, MST)	Mesotocin (MT, MST)	Mesotocin (MT) Isotocin (IT, IST) Glumitocin, Neurophysin, IT-1 like, IT-NP	Valitocin Aspargtocin	**Oxytocin** **OT**
OXTR, OTR	VT3, MTR	OXTR	MesoR, OXTR	ITR, OXTR, itnpr-like 2, itr2	OXTR	**Oxytocin Receptor** **OTR**
Arginine Vasopressin (AVP, ARVP, AVRP, Vp, Vsp) Neurophysin II (NPII) Lysine vasopressin Phenypresin	Vasotocin (VT)	Vasotocin (VT)	Vasotocin (VT)	Vasotocin (VT) VT-NP, avpl, vsnp	Vasotocin (VT)	**Vasotocin** **VT**
AVPR1a, V1aR, V1A	VT4, VT4R		Avpr1, VasR	Avpr1aa, VasR, Avpr1ab		**Vasotocin Receptor 1A** **VTR1A/V1A**
AVPR1b, V1bR, AVPR3, V3, VIBR, VPR3	VT2, AVT2R					**Vasotocin Receptor 1B** **VTR1B/V1B**
AVPR2, V2R, VPV2R				Avpr2bb, V2A(2), avpr2a, AVPR2A.A		**Vasotocin Receptor 2A** **VTR2A/V2A**
				V2B, V2BR1, V2Rl, OTRl, nft, avpr2		**Vasotocin Receptor 2B** **VTR2B/V2B**
	VT1, AVPR2		Avpr2.2	V2C, V2bR2, Avpr2.2, V2L	V2C, V2bR2	**Vasotocin Receptor 2C** **VTR2C/V2C**

Table 2: Aliases for all oxytocin and vasopressin/vasotocin ligands and receptors in major vertebrate lineages used in the literature, and our proposed universal vertebrate revision (last column). Long (e.g. VTR1A) and short (e.g. V1A) versions for the gene names we propose are given in the last column.

	Organism	OT	VT	OTR	VTR1A	VTR1B	VTR2A	VTR2B	VTR2C
Mammals	Human (*Homo Sapiens*)								
	Chimpanzee (*Pan troglodytes*)								
	Western Gorilla (*Gorilla gorilla*)								
	Northern white-cheeked Gibbon (*Nomascus leucogenys*)								
	Rhesus Macaque (*Macaca mulatta*)								
	Marmoset (*Callithrix jacchus*)								
	Mouse lemur (*Microcebus murinus*)								
	Mouse (*Mus musculus*)								
	Prarie Vole (*Microtus ochrogaster*)								
	Cow (*Bos taurus*)								
	Yangtze River Dolphin (*Lipotes vexillifer*)								
	Horse (*Equus caballus*)								
	Dog (*Canis lupus familiaris*)								
	Megabat (*Pteropus vampyrus*)								
	Platypus (*Ornithorhynchus anatinus*)								
Birds	Chicken (*Gallus gallus*)								
	Anna's hummingbird (*Calypte anna*)								
	Zebra finch (*Taeniopygia guttata*)								
Reptiles	American Alligator (*Alligator mississippiensis*)								
	Carolina anole-lizard (*Anolis carolinensis*)								
	Painted turtle (*Chrysemys picta bellii*)								
Amphibia	Tropical clawed frog (*Xenopus tropicalis*)								
Holost fish	Spotted Gar (*Lepisosteus oculatus*)								
Teleost fish	Southern platyfish (*Xiphophorus maculatus*)								
	Three-spined stickleback (*Gasterosteus aculeatus*)								
	Japanese medaka (*Oryzias latipes*)								
	Nile Tilapia (*Oreochromis niloticus*)								
	Zebrafish (*Danio rerio*)								
Coelacanths	Coelacanth (*Latimeria chalumnae*)								
Sharks	Elephant Shark (*Callorhinchus milii*)								
Lampreys	Japanese lamprey (*Lethenteron japonicum*)								
	Sea Lamprey (*Petromyzon marinus*)								

Present
Only the second copy is present
Deleted
Not present
Sequencing gap

Table 3: Presence versus absence of orthologs of nonapeptide genes and their receptors based on microsynteny analysis across all vertebrate lineages. Microsynteny results for each gene are presented in Tables S1a-S1h. To be considered orthologous, the focus gene had to be syntenic with 10 genes, 5 on either side, in at least one species with a well assembled genome of the specific vertebrate lineage.

Figures

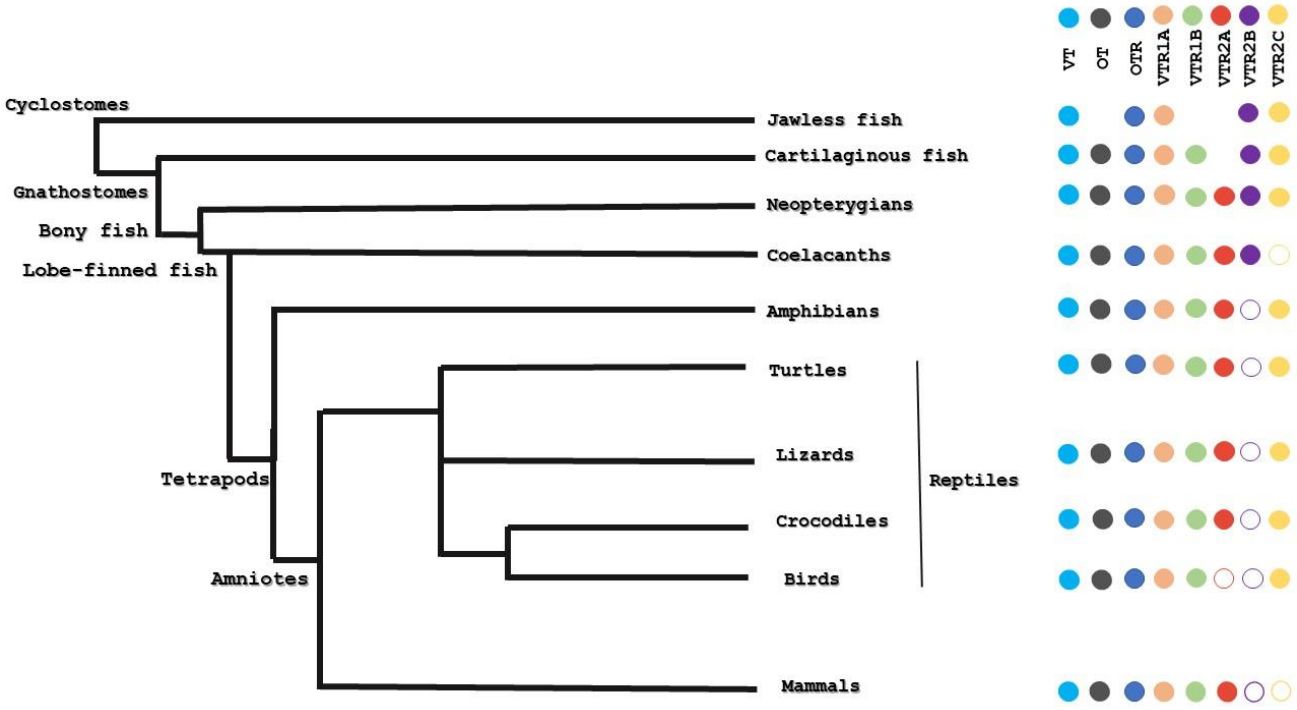


Figure 1: Distribution of OT-VT and OTR-VTR in all major vertebrate lineages, using our revised universal nomenclature. Full colored circles denote presence of gene; empty circles denote deletion of gene; no circle denotes the gene never evolved in that lineage.

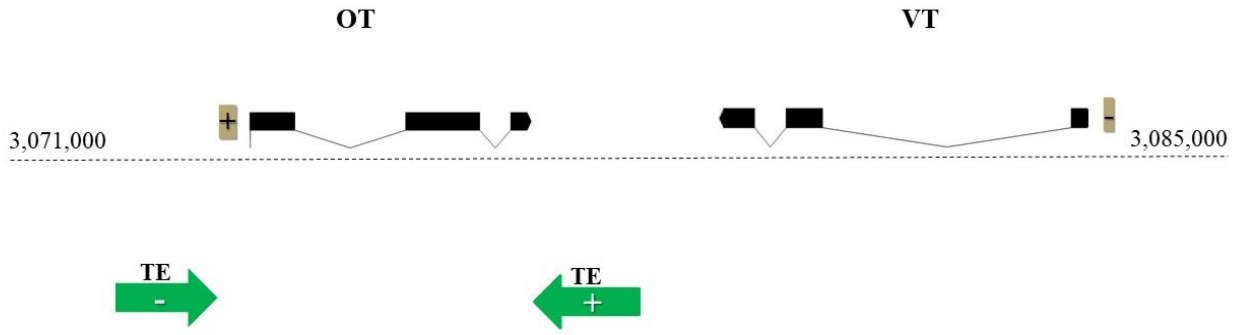


Figure 2: Representation of the position (kb) and orientation (+, -) of OT and VT genes (grey) and DNA transposable elements (green) next to OT in the human genome (chromosome 12). Exon-intron structure for OT and VT is shown to illustrate intron shortening in OT compared to VT (scale length: 100 bases).

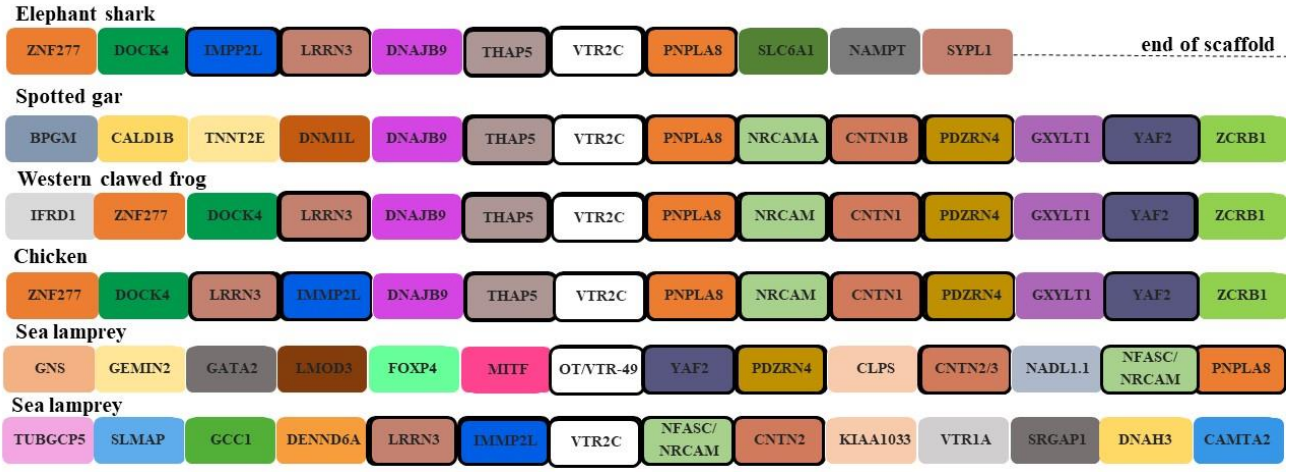


Figure 3: Synteny summary for VTR2C territory for a shark (elephant shark), fish (spotted gar), amphibian (western clawed frog), bird (chicken), and sea lamprey. In the sea lamprey, OT/VTR-49 is our revised nomenclature for PMZ_0045207-RA on scaffold_49 (Table S4 for specific location), and VTR2Cis our revision for PMZ_0042163-RA on scaffold_10, the putative VTR2C-ortholog. Orthologous genes are colored with the same color and genes that are found in the territory of the sea lamprey genes are further contoured.

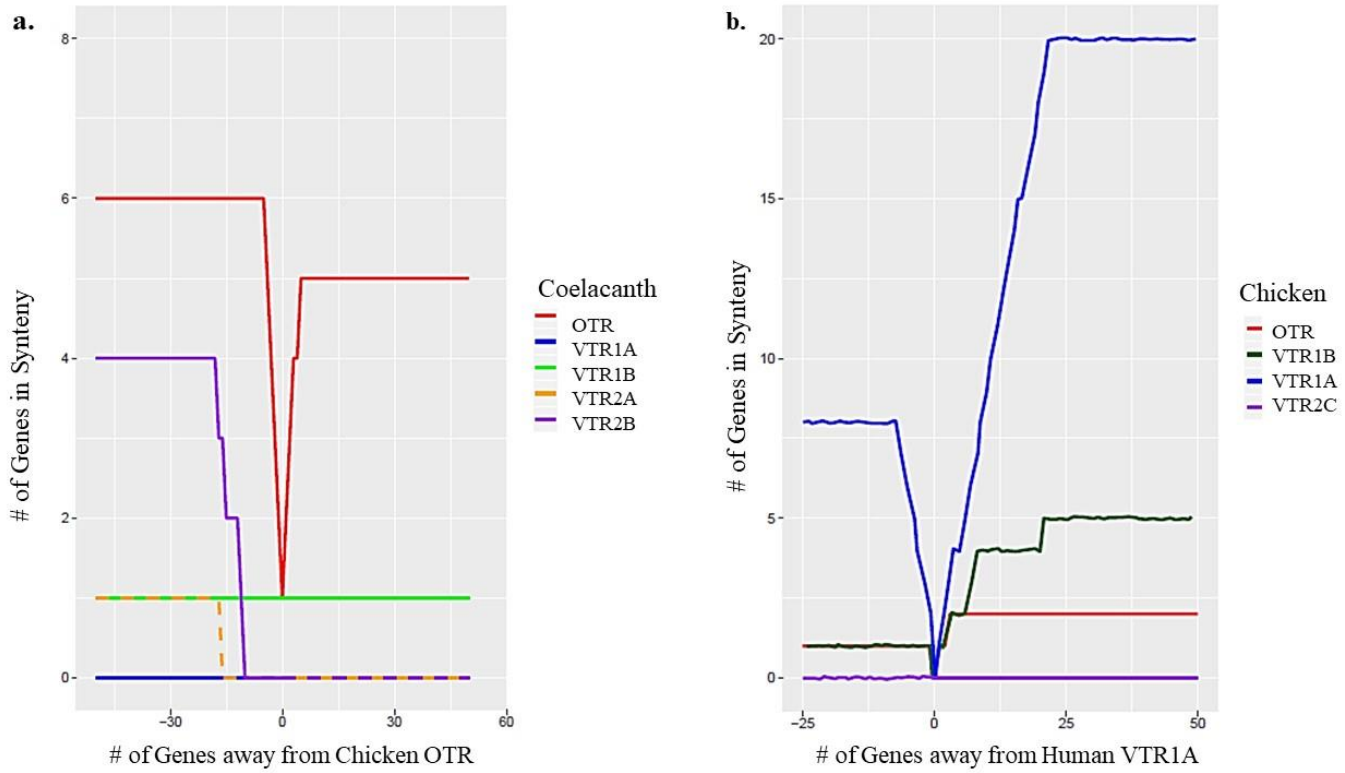


Figure 4: Plots generated with SynFind. In the x axis, 0 represents the query OTR or VTR receptor in the query organism (e.g. OTR in chicken in a.) and the numbers represent the genes on the 5' (left) and on the 3' (right) of the query OTR or VTR in the genome (e.g. '30' genes on the left of '0' in a. represents the 30th gene on the left of OTR in the chicken genome). The y axis shows the number of matched homologous genes in the reference genome for each reference receptor (e.g. in a. coelacanth OTR in red shows 6 syntenic gene matches with genes on the left of chicken OTR, and 5 matches on the right of chicken OTR). If the reference OTR or VTR does not show any match, then it stays 0 in the y axis (e.g. in a. coelacanth VTR1A, in blue); if it matches only the query OTR-VTR, it reaches 1 (e.g. in a. coelacanth VTR1B -in green- was only homologous to chicken OTR). If the reference OTRVTR is not homologous to the query OTR-VTR but does show gene matches in the neighboring territory, then its line increases where the gene-match is located (e.g. in a. coelacanth VTR2A -in yellow- hit a gene match that is found approximately 15 genes on the left of chicken OTR, which is where its line increased and reached 1 in the y axis).

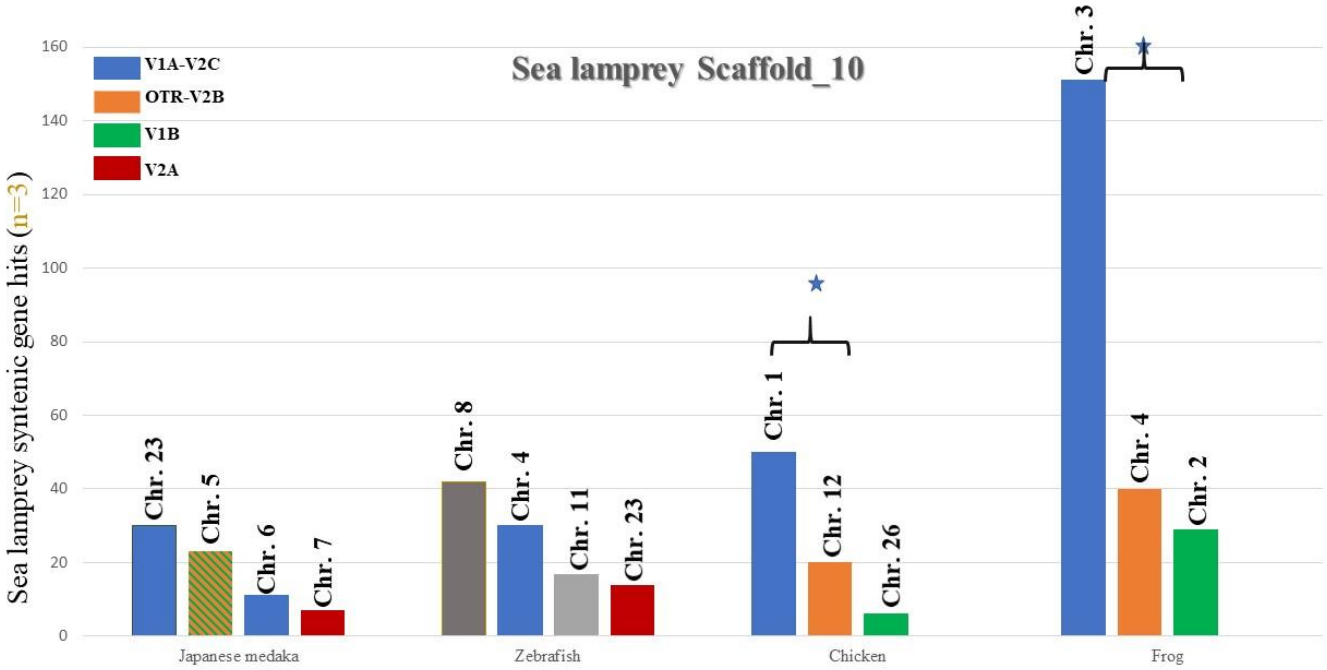


Figure 5: Number of syntenic gene hits between Japanese medaka, zebrafish, chicken and frog chromosomes against sea lamprey scaffold_10. * indicate that the difference in gene hits between the first two chromosomes with the highest number of hits is statistically different ($p<0.05$; t test). n=3 for lamprey (y-axis) means that the minimum number of aligned homologous gene pairs (at a maximum of a 20-gene distance from each other in each genome) for these genes to be considered syntenic was 3. Chromosomes are colored based on the OTR-VTR that reside in them.

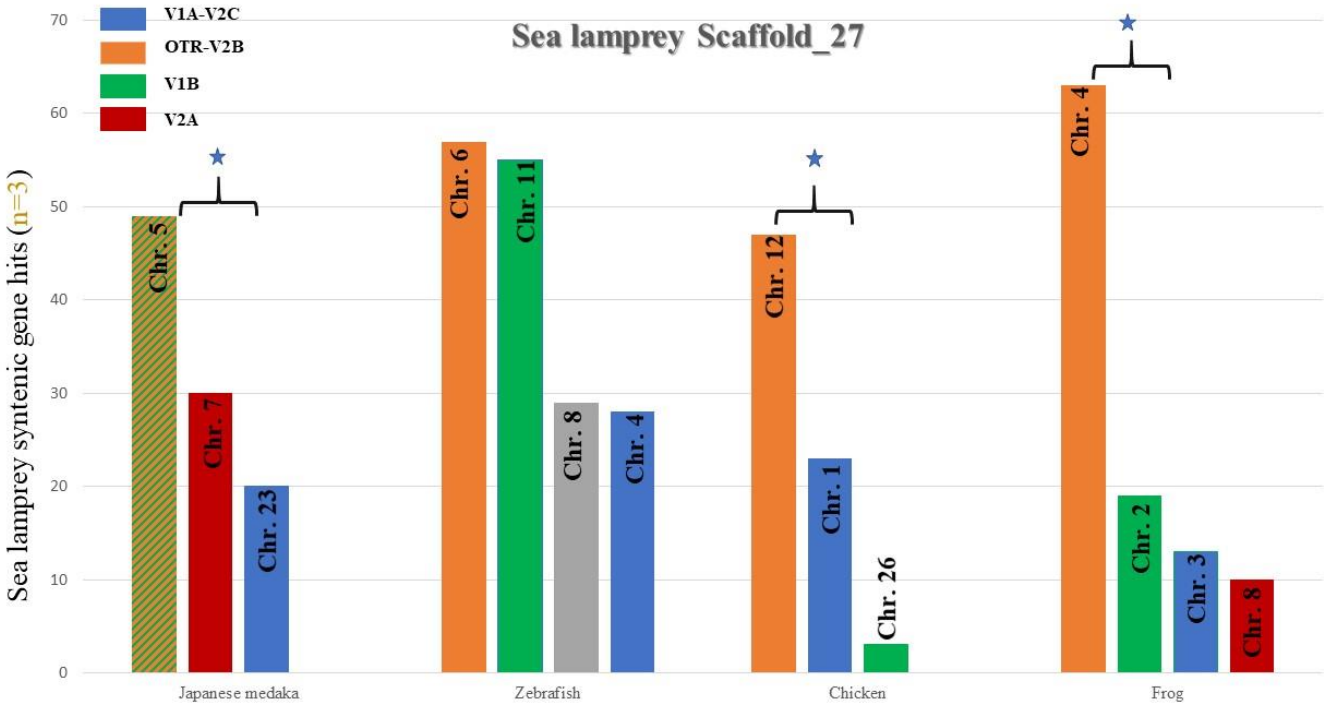


Figure 6: Number of syntenic gene hits between Japanese medaka, zebrafish, chicken and frog chromosomes against sea lamprey scaffold_27. * indicate that the difference in gene hits between the first two chromosomes

with the highest number of hits is statistically different ($p<0.05$; t test). n=3 means that the minimum number of aligned homologous gene pairs (at a maximum of a 20-gene distance from each other in each genome) for these genes to be considered syntenic was 3. Chromosomes are colored based on the OTR-VTR that reside in them.

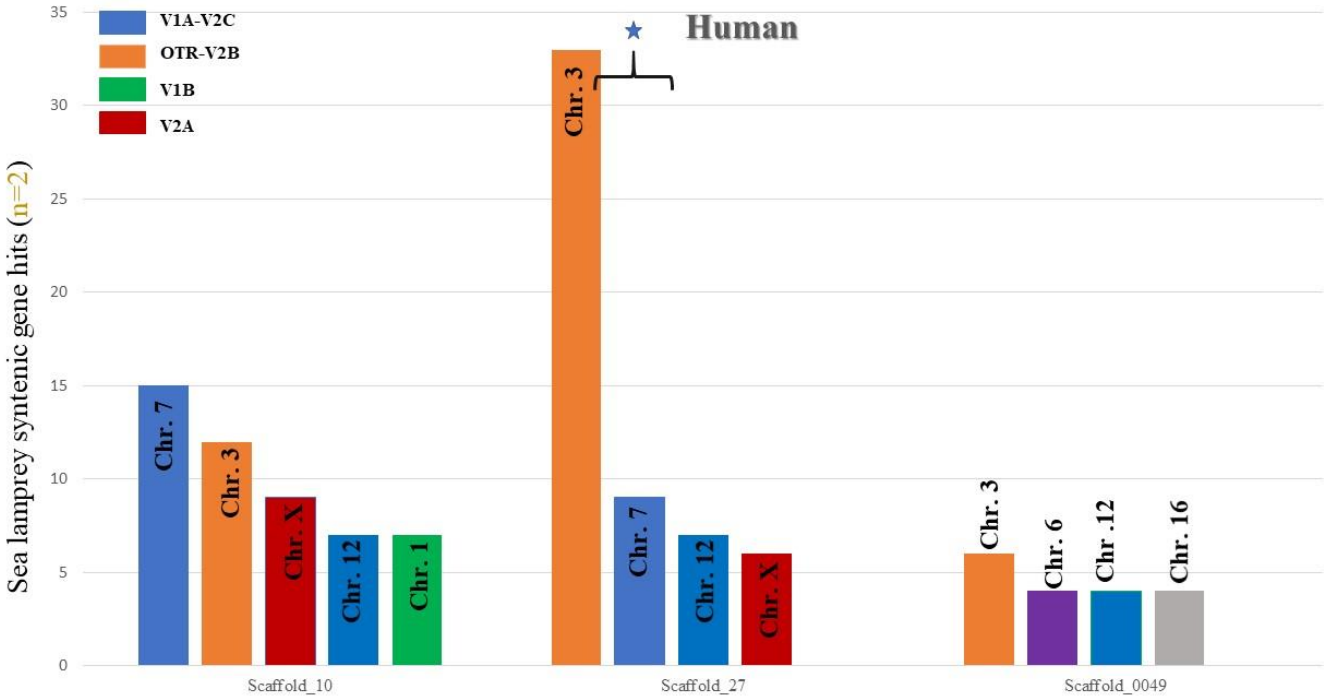
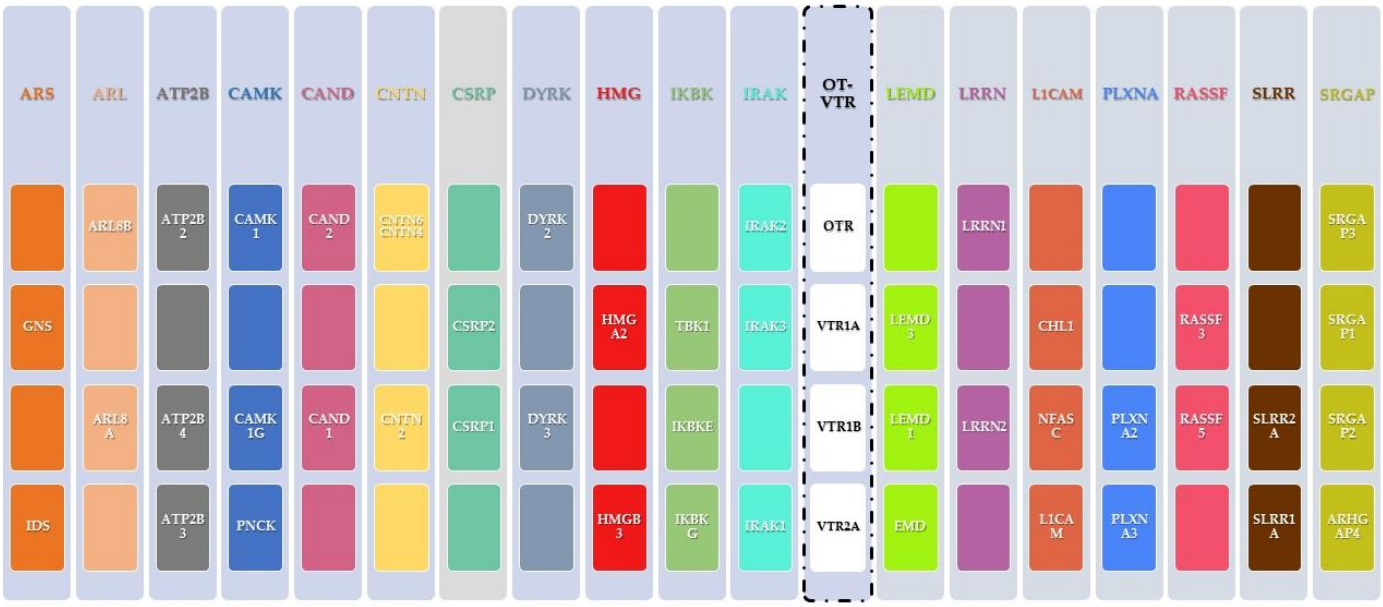


Figure 7: Number of syntenic gene hits between sea lamprey scaffolds 10, 27 and 49 against the human chromosomes. * indicate that the difference in gene hits between the first two chromosomes with the highest number of hits is statistically different ($p<0.05$; t test). n=2 means that the minimum number of aligned homologous gene pairs (at maximum of a 20-gene distance from each other in each genome) for these genes to be considered syntenic was 2. Chromosomes are colored based on the OTR-VTR that reside in them.



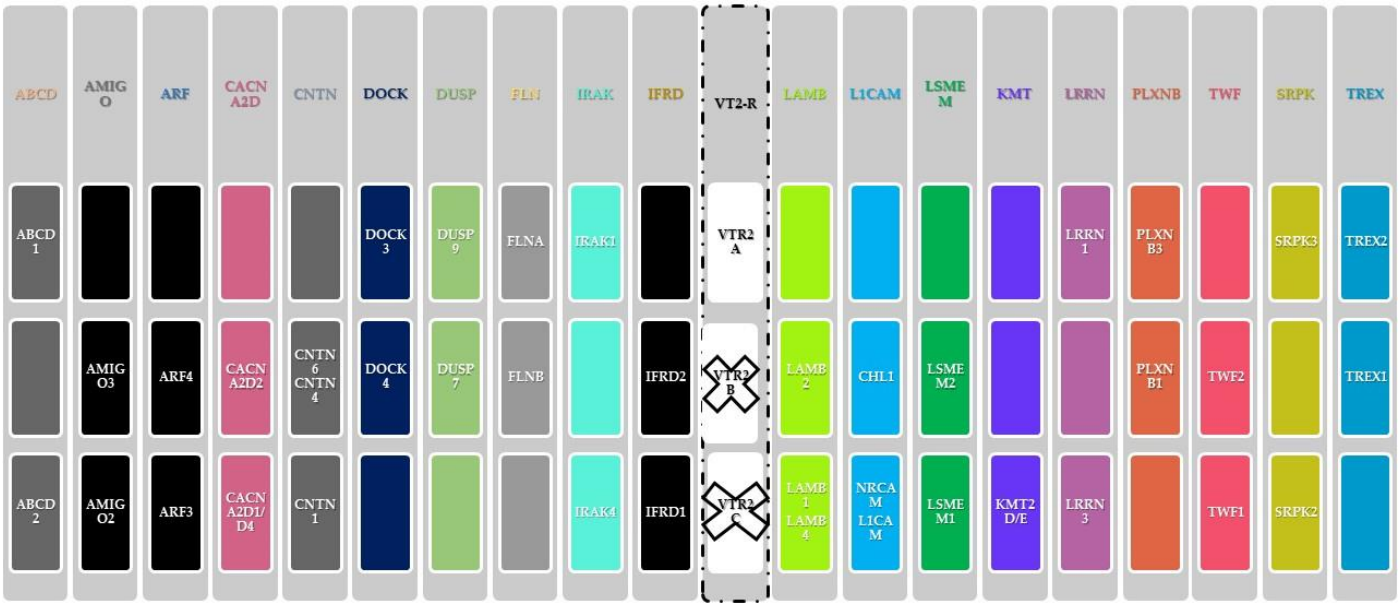


Figure 8: Between chromosome intraspecies synteny for the OTR-VTRs in humans. Top) Gene families in a 10 Mb window surrounding the four OTR-VTRs present in the human genome. Bottom) Gene families in a 10 Mb or entire chromosomal window surrounding the three VTR2A and proposed paralogous missing VTR2s in the human genome (VTR2B, VTR2C). The window size was made bigger for this analysis, due to apparent deleted genomic DNA around the missing genes. For gene family-symbols, we followed the ones proposed by the HGNC (HUGO Gene Nomenclature Committee).

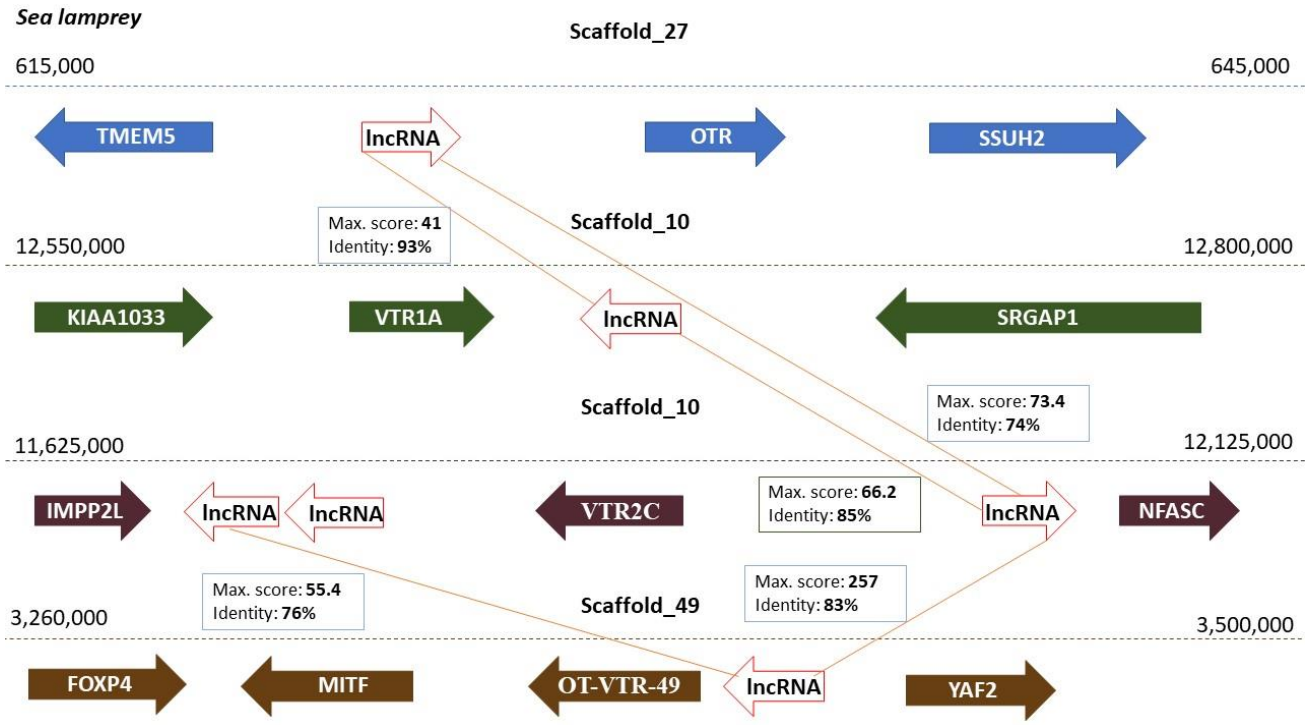


Figure 9: Long non-coding RNAs (lncRNAs) around the OTR and VTRs in sea lamprey. Lines connect the lncRNAs that shared identity beyond threshold (maximum score>40 and E-value< 10^{-4}) in the BLASTn

comparisons. Max. score (bitscore) and percent identity are shown for each pair of lncRNAs. Genomic location is in kb.

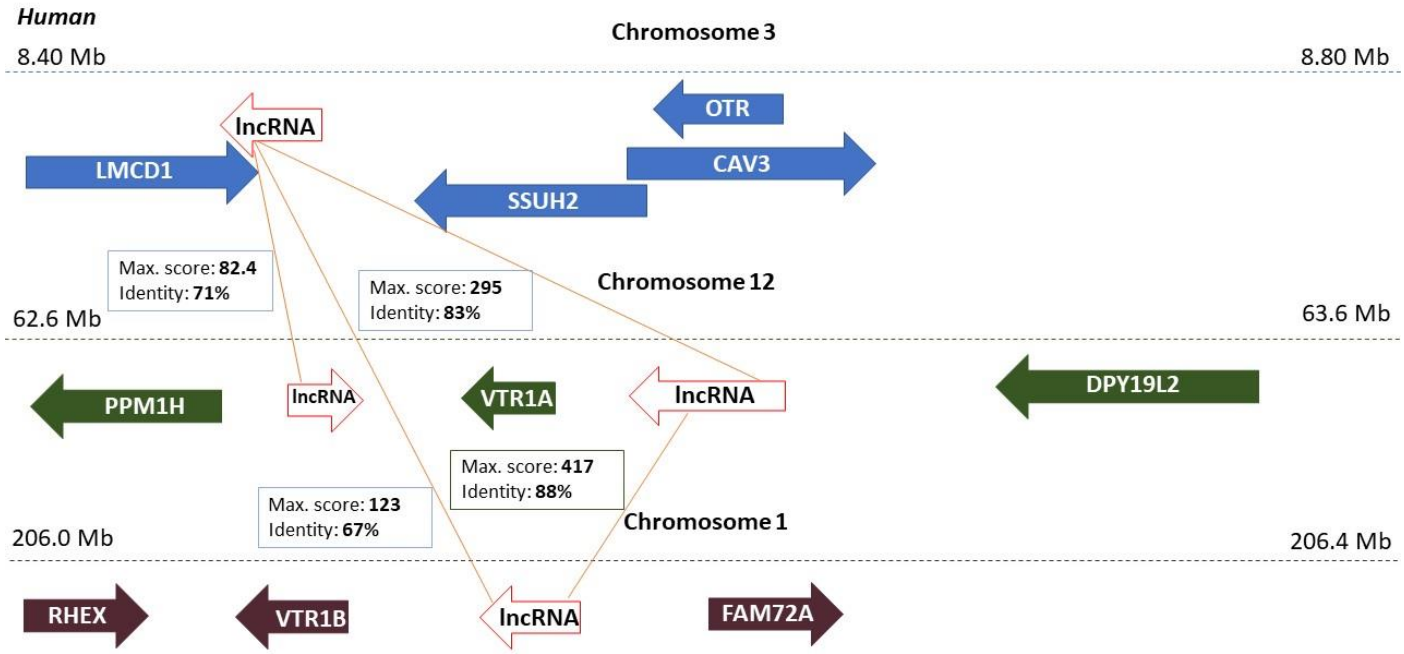


Figure 10: Long non-coding RNAs (lncRNAs) around the OTR and VTR1s in human. Lines connect the lncRNAs that shared identity beyond threshold (maximum score>40 and E-value< 10^{-4}) in the BLASTn comparisons. Max. score (bitscore) and percent identity are shown for each pair of lncRNAs. Genomic location is in Mb.

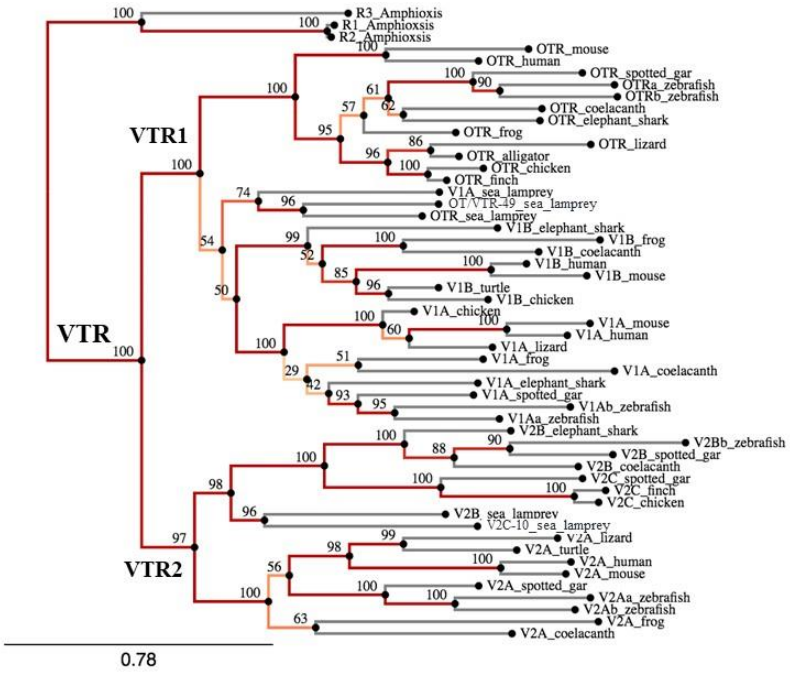


Figure 11: Phylogenetic relationships between the OTR, VTR1A/V1A, VTR1B/V1B, VTR2A/V2A, VTR2B/V2B, VTR2C/V2C are shown. Tree topology was inferred with the phylogenetic maximum likelihood method from an exonic sequence alignment (MAFFT), supported by a non-parametric bootstrap analysis with 100 replicates. Bootstrap values are shown as percentages at the branch points (values <50% are not considered informative). The tree is rooted with the 3 putative OTR-VTRs we found in amphioxus (Accession IDs: 154074, 154241, 134295). The scale bar indicates phylogenetic distance of 0.78 substitutions per site. Only sequences with 100bp or more of assembled exons were included to prevent tree inference artifacts; this resulted in not including VTR2C for all three reptiles and frog, due to incomplete assemblies. Full scientific and common names of organisms and gene accession numbers are included in Table S1.

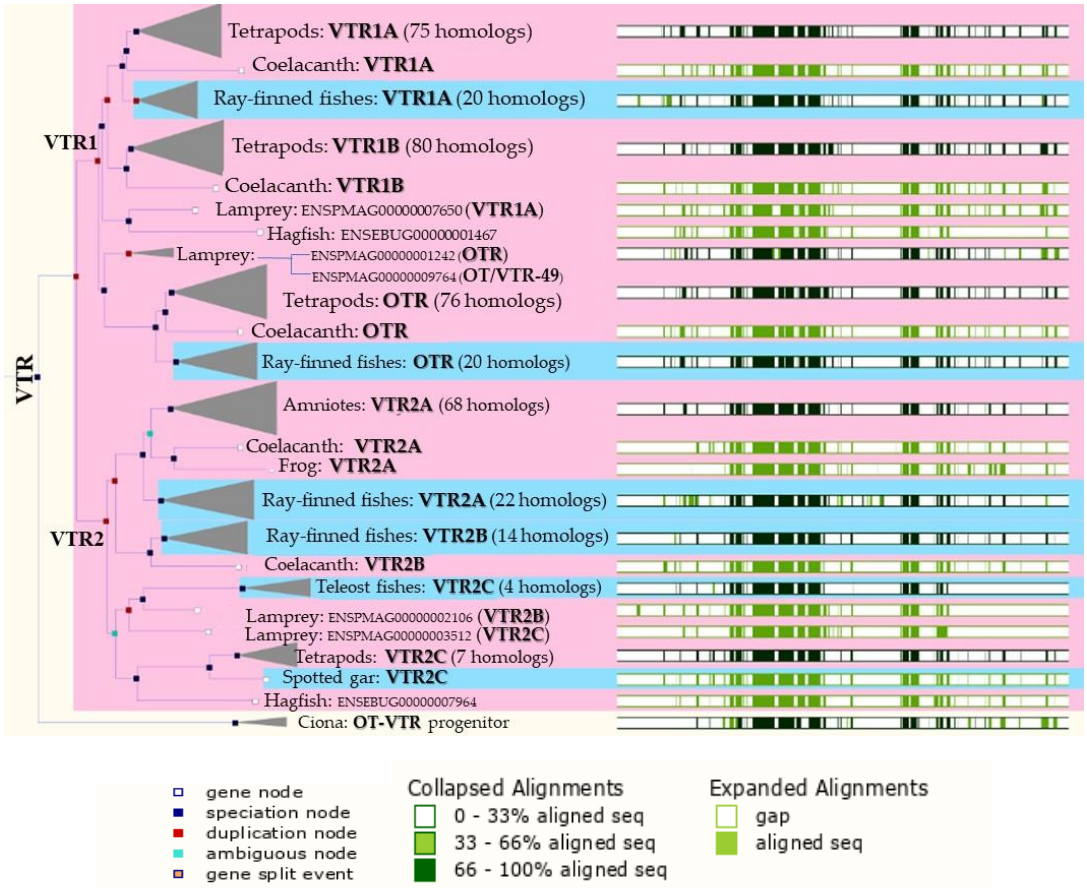


Figure 12: Maximum likelihood phylogenetic tree generated via the Ensembl 'Gene tree' tool (Gene Tree ID: ENSGT00760000119156) that uses the Gene Orthology/Paralogy prediction method pipeline. The phylogeny was constructed using one representative protein for every gene in every species in Ensembl (the longest available). The tree is reconciled with a species tree, generated by TreeBeST. Internal nodes are annotated for duplication (red boxes) or speciation (blue boxes) events. Multiple alignment of the peptides (green bars) was made with MUSCLE. Green bars show areas of amino acid alignment, white areas are gaps in the alignment. Dark green bars indicate consensus alignments. Bootstrap support values that are very low are highlighted in turquoise boxes, annotated as 'ambiguous nodes' by Ensembl. We curated the Ensembl tree and re-named genes using the universal, more syntenic-based nomenclature we proposed in this study.

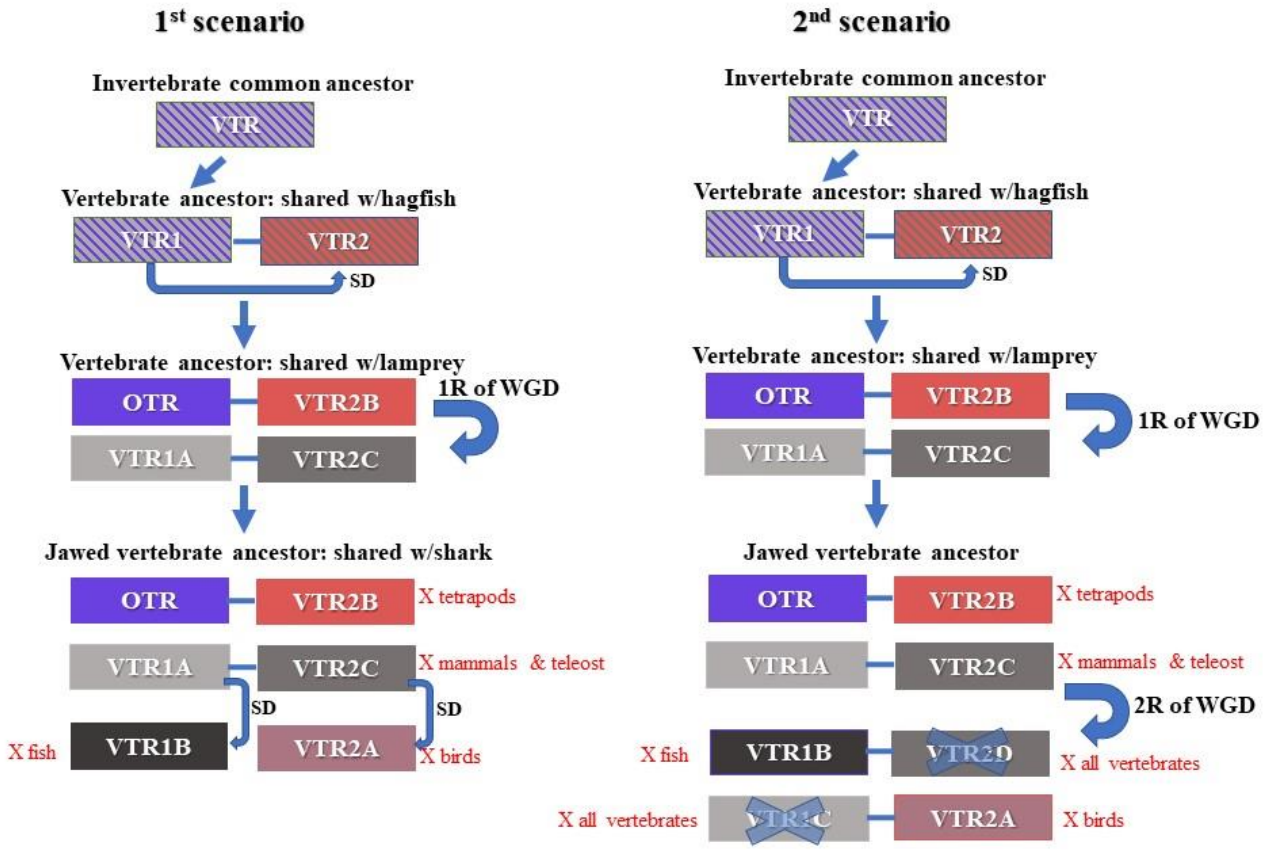


Figure 14: Two proposed scenaria for the evolution of OTR-VTRs. According to both scenaria, the vertebrate ancestor (represented by shared organization between hagfish and other vertebrates) carried a VTR1 progenitor gene that gave rise to a VTR2 progenitor gene resulting from a tandem duplication. One round of whole-genome duplication (1R WGD) created two syntenic blocks, one containing VTR1A-VTR2C (on scaffold 10 in sea lamprey) and the other OTR-VTR2B (on scaffold 27 in sea lamprey). There was a species- or order-specific segmental duplication that gave rise to PMZ_0014716-RA and PMZ_0045207-RA on scaffold 49, most likely a duplication from scaffold 27. Thereafter, in the 1st scenario two segmental duplications occurred that created VTR1B and VTR2A: the first segmental duplication happened after the split of gnathostomes from cyclostomes and created VTR1B from VTR1A; the second segmental duplication happened after the split of bony fish from gnathostomes and created VTR2A from VTR2C. According to the 2nd scenario, there was a second round of WGD after the divergence of gnathostomes from cyclostomes that formed four syntenic blocks in jawed vertebrates. In this scenario, 2 gene losses in individual all lineages in hypothetical lineages at the origin of vertebrates have to be assumed.

Supplementary Material

Supplementary Tables

Organism	Chromosome/ Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Adjacent genes
Human (Homo Sapiens)	20p13	NC_000020.11	3082555-3093521	ENSG00000101200	4	VPS16, PTPRA,GNRH2,MRPS26,LOC, OXT,* ,UBOX5,FASTKD5,LZTS3,LOC, DDRGK1,ITPA
Chimpanzee (Pan troglodytes)	20	NC_006487.4	2938436-2941286	ENSPTRG00000013190	3	VPS16, PTPRA,GNRH2,MRPS26,OXT,*, UBOX5,FASTKD5,LOC, DDRGK1,ITPA, SLC4A11
Western Gorilla (Gorilla gorilla)	20	NC_018444.2	2961383-2964188	ENSGGOG00000027655.2	3	VPS16, PTPRA,GNRH2,MRPS26, OXT,*,UBOX5,FASTKD5,LZTS3, LOC, DDRGK1,ITPA
Gibbon (Nomascus leucogenys)	13	NC_019828.1	36650325-36653605	ENSNLEG00000007575	3	VPS16, PTPRA,GNRH2,MRPS26, OXT, *,UBOX5,FASTKD5,LOC, DDRGK1,ITPA, SLC4A11
Rhesus Macaque (Macaca mulatta)	10	NC_027902.1	35910800-35922951	ENSMMUG00000041847	4	VPS16, PTPRA,GNRH2,MRPS26, OXT, *, UBOX5,FASTKD5,PROSAPIP1,DDRGK1,ITPA
Marmoset (Callithrix jacchus)	5	NC_013900.1	8875483-8878232	ENSCJAG00000021094	3	VPS16, LOC, PTPRA,LOC, LOC, GNRH2,MRPS26,OXT, *, LOC, UBOX5,FASTKD5,LZTS3,DDRGK1, ITPA
Mouse lemur (Microcebus murinus)	18	NC_033677.1	22782854-22787821	ENSMICG00000026411	3	VPS16, PTPRA,GNRH2,MRPS26, OXT, *, UBOX5,FASTKD5,LZTS3, DDRGK1,LOC, ITPA
Mouse (Mus musculus)	2	NC_000068.7	130580620-130582588	ENSMUSG00000037727	3	VPS16, PTPRA,GNRH2,4930473A02Rik, MRPS26, OXT, *, UBOX5,FASTKD5,LZTS3, DDRGK1,ITPA
Prarie Vole (Microtus ochrogaster)	JH996433.1	NC_004949101.1	17192904-17194907	ENSMOCG00000014821	3	PCED1A, VPS16, PTPRA, MRPS26, OXT, *, UBOX5,FASTKD5,LOC, DDRGK1,ITPA, SLC4A11
Cow (Bos taurus)	13	AC_000170.1	52563657-52565676	ENSBTAG00000008027	3	ITPA, DDRGK1,PROSAPIP1,FASTKD5, UBOX5, *, OXT, MRPS26, GNRH2, PTPRA, VPS16
Yangtze River Dolphin (Lipotes vexillifer)	1300	NW_006792177.1	2086597-2088604	103076585	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26, OXT, *, UBOX5, FASTKD5, LOC, DDRGK1, ITPA, SLC4A11
Horse (Equus caballus)	22	NC_009165.2	19691102-19693167	100066842	3	SLC4A11, ITPA, DDRGK1,LOC, FASTKD5, UBOX5, *, OXT (ENSECAG00000007628), MRPS26, GNRH2, PTPRA, VPS16
Dog (Canis lupus familiaris)	24	NC_006606.3	18183057-18184827	ENSCAFG00000006437	3	ITPA, DDRGK1,LOC, LZTS3, FASTKD5, UBOX5, *, OXT, MRPS26, GNRH2, PTPRA, VPS16
Megabat (Pteropus vampyrus)	629	not available	823-2761	ENSPVAG00000014205	10	*, UBOX5, FASTKD5, LZTS3, DDRGK1, ITPA
Platypus (Ornithorhynchus anatinus)	18368	NW_001655452.1	1060-5789	100089405	3	*,OXT,MRPS26
Chicken (Gallus gallus)	4	NC_006091.4	89222770-89224343	ENSGALG00000014117	3	PANK2, MAVS, PTPRA, MRPS26, OXT, *, UBOX5, FASTKD5, LZTS3, LOC, DDRGK1, HTR7
Anna's hummingbird (Calypte anna)	Unk	NW_007619440.1	960073-968915	103526080	4	SLC4A11, DDRGK1, LOC, FASTDK5, UBOX5, *, MRPS26, PTPRA
Zebra finch (Taeniopygia guttata)	4	NC_011467.1	69076722-69078936	ENSTGUG00000010982	3	PANK2, MAVS, PTPRA, LOC, MRPS26, OXT, *, UBOX5, FASTKD5, LOC, DDRGK1, LOC, SLC4A11
Green anole-lizard (Anolis carolinensis)				LIKELY PRESENT		scaffolding gap
American Alligator (Alligator mississippiensis)	Unk	NW_017712067.1	5452969-5457507	102569046	3	MAVS, PTPRA, GNRH2, MRPS26, OXT, *, LOC, UBOX5, FASTKD5, LOC, DDRGK1, LOC, LOC,SLC4A11
Painted turtle (Chrysemys picta bellii)	Unk	NW_007281366.1	6047893-6054528	101938631	3	SLC4A11, LOC, DDRGK1, LOC, FASTKD5, UBOX5, *, OXT, MRPS26, GNRH2, LOC, PTPRA, MAVS
Three-spined stickleback (Gasterosteus aculeatus)	XII		5162684-5164321	ENSGACG00000006670	3	OXT, FABP1B.1, SMYD1B, LZTSE3B, UBOX5, *, SLC35D2, ZNF367, HABP4, CDC14B, AAED1
Tropical clawed frog (Xenopus tropicalis)	1	NC_030677.1	6321594-6327868	100038191	3	PANK2, MAVS, PTPRA, GNRH2, OXT, *, UBOX5, FASTKD5, LZST3, LOC, GFRA4, ATRN
Southern platyfish (Xiphophorus maculatus)	JH556662.1	NW_005372180.1	3963904-3966467	ENSXMAG00000016007	3	OXT, FABP1B.1, SMYD1B, LZTSE3B, UBOX5, *, SLC35D2, ZNF367, HABP4, CDC14B, AAED1
Spotted Gar (Lepisosteus oculatus)	LG2	NC_023180.1	441340-445005	ENSLOCG00000001161	3	PTPRA, IDH3B,IL12B,DQX1, *, OXT, THNSL2,FABP1A,SMYD1B, SPRA, DDRGK1,LZST3B
Japanese medaka (Oryzias latipes)	9	NP_001265820	7013517-7014827	ENSORLG00000003907	3	OXT , FABP1B.1, SMYD1B, LZST3B, UBOX5, *, SLC35D2, ZNF367, HABP4, CDC14B, AAED1
Nile Tilapia (Oreochromis niloticus)	LG12	NC_022210.1	21545532-21547222	ENSONIG00000015218	3	AAED1, CDC14B, HABP4, ZNF367, SLC35D2, *, UBOX5, LZST3B, SMYD1B, FABP1B.1, OXT
Zebrafish (Danio rerio)	8	NC_007119.7	1148876-1157380	ENSDARG00000058567	3	FABP1B.2, SMYD1B, SPRB,LZTS3B,UBOX5,*,CCL27A,SLC35D2,ZNF367 CC14B, AAED1
Coelacanth (Latimeria chalumnae)	Unk	NW_005820294.1	504462-514162	ENSLACG00000019419	3	MAVS,PTPRA,GNRH2, OXT, *,UBOX5,FASTKD5, LZTS3
Elephant Shark (Callorhinchus milii)	Unk	NW_006890112.1	2505593-2512045	102181495	5	DDRGK1, LOC, FASTKD5, UBOX5, *, OXT, PTPRA, MAVS, PANK2, RNF24, SMOX
Japanese Lamprey (Lethenteron camtschaticum)	scaffold00009			KE993680.1		
Sea Lamprey (Petromyzon marinus)	scaf_00015	PMZ_0041499-RA	5718006-5724473	not available	3	TMEM180, NOP56, LSM11, COE3, EBF3,*, PTPRA, FAM45A, EIF3A, NANOS1, LOC, PRLHR

Table S1a: Microsynteny analysis (10-gene window) for VT.

Organism	Chromosome /Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Adjacent genes
Human (Homo Sapiens)	20p13	NC_000020.11	3068871-3072517	ENSG00000101200	4	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,LOC, *,AVP,UBOX5,FASTKD5,LZTS3,LOC, DDRGK1
Chimpanzee (Pan troglodytes)	20	NC_006487.4	2927405-2928332	ENSPTRG00000013189	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,*, AVP,UBOX5,FASTKD5, LOC, DDRGK1,ITPA
Western Gorilla (Gorilla gorilla)	20	NC_018444.1	2936881-2951225	ENSGGOG00000011857	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,*, AVP,UBOX5,FASTKD5, LZTS3, LOC, DDRGK1
Gibbon (Nomascus leucogenys)	13	NC_019828.1	36638388-36639275	ENSNLEG00000007574	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,*, AVP,UBOX5,FASTKD5,LOC, DDRGK1, ITPA
Rhesus Macaque (Macaca mulatta)	10	NC_027902.1	35900674-35901743	ENSMMUG00000010678	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,*, AVP,UBOX5,FASTKD5,PROSAPIP1, DDRGK1
Marmoset (Callithrix jacchus)	5	NC_013900.1	8860308-8861216	ENSCJAG00000021091	3	PCED1A, VPS16, LOC, PTPRA,LOC, LOC, GNRH2,MRPS26,*, AVP, LOC, UBOX5,FASTKD5,LZTS3, DDRGK1
Mouse lemur (Microcebus murinus)	18	NC_033677.1	22757696-22761910	ENSMICG00000035095	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,*, AVP,UBOX5,FASTKD5,LZTS3, DDRGK1
Mouse (Mus musculus)	2	NC_000068.7	130574520-130577054	ENSMUSG00000027301	3	PCED1A, VPS16, PTPRA,GNRH2,4930473A02Rik, MRPS26,*, AVP,UBOX5,FASTKD5, LZTS3, DDRGK1
Prarie Vole (Microtus ochrogaster)	JH996433.1	NC_004949101.1	17198627-17199468	ENSMOCG00000017703	3	TMEM239, PCED1A, VPS16, PTPRA, MRPS26,*, AVP,UBOX5,FASTKD5,LOC, DDRGK1,ITPA
Cow (Bos taurus)	13	AC_000170.1	52575290-52578188	ENSBTAG00000008026	3	DDRGK1, PROSAPIP1, FASTKD5, UBOX5, AVP, *, MRPS26, GNRH2, PTPRA, VPS16, PCED1A
Yangtze River Dolphin (Lipotes vexillifer)	1300	NW_006792177.1	2073772-2074672	ENSTTRG00000011964	3	PCED1A, VPS16, PTPRA,GNRH2,MRPS26,*, AVP,UBOX5,FASTKD5, LOC, DDRGK1
Horse (Equus caballus)	22	NC_009165.2	19701308-19701748	ENSECAG00000007628	3	ITPA, DDRGK1, LOC, FASTKD5, UBOX5, AVP, *, MRPS26, GNRH2, PTPRA, VPS16, PCED1A
Dog (Canis lupus familiaris)	24	NC_006606.3	18193381-18194231	ENSCAFG00000006432	3	DDRGK1, LOC, LZTS3, FASTKD5, UBOX5, AVP, *, MRPS26, GNRH2, PTPRA, VPS16, PCED1A
Megabat (Pteropus vampyrus)	90445	not available	4-578	ENSPVAG00000000327.1	2	only gene on scaffold
Platypus (Ornithorhynchus anatinus)	18368	NW_001655452.1	11193-12510	ENSOANT00000020606	3	AVP, *, MRSPS26
Chicken (Gallus gallus)	4	NC_006091.4	89222770-89224343	768516	3	RNF24, PANK2, MAVS, PTPRA, MRPS26, *, AVP, UBOX5,FASTKD5, LZTS3, LOC, DDRGK1
Anna's hummingbird (Calypte anna)	scaffold82	NW_007619440.1	960073-968915)	103526080	4	LOC,FASTKD5,UBOX5,*,MRPS26,PTPRA
Zebra finch (Taeniopygia guttata)	4	NC_011467.1	69072345-69073893	ENSTGUG00000010980	3	RNF24, PANK2, MAVS, PTPRA, LOC, MRPS26, *, AVP, UBOX5, FASTKD5, LOC, DDRGK1
Green anole-lizard (Anolis carolinensis)				LIKELY PRESENT		scaffolding gap
American Alligator (Alligator mississippiensis)	Unk	NW_017712067.1	5434824-5441858	102569278	3	PANK2, MAVS, PTPRA, GNRH2, MRPS26, *, AVP, LOC, UBOX5, FASTKD5, LOC, DDRGK1, LOC, LOC,SLC4A11
Painted turtle (Chrysemys picta bellii)	Unk	NW_007281366.1	6069639-6074797	101938903	3	SLC4A11, LOC, DDRGK1, LOC, FASTKD5, UBOX5, AVP, *, MRPS26, GNRH2, LOC, PTPRA, MAVS, PANK2
Three-spined stickleback (Gasterosteus aculeatus)	XIII	not available	5131577-5133164	ENSGACG00000006569	4	RNF165B, LOXHD1B, PTGER4A, MRPS27, ZNF366, *, FABP1B.1, SMYD1B, LZTSE3B, UBOX5, AVP
Tropical clawed frog (Xenopus tropicalis)	1	NC_030677.1	6307623-6312315	100485685	3	RNF24, PANK2, MAVS, PTPRA, GNRH2, *, AVP, UBOX5, FASTKD5, LZST3, LOC, GFRA4
Southern platyfish (Xiphophorus maculatus)	JH556662.1	NW_005372180.1	4015717-4018740	ENSXMAG00000016093	4	AVP, UBOX5, LZST3B, SMYD1B, FABP1B.1, *, ZNF366, MRPS27, PTGER4A, LOXHD1B, MAPKAP5
Spotted Gar (Lepisosteus oculatus)	LG2	NC_023180.1	432706-437919	ENSLOCG00000001173	3	PTPRA, IDH3B,IL12B,DQX1, AVP, *, THNSL2,FABP1A,SMYD1B, SPRA, DDRGK1,LZST3B
Japanese medaka (Oryzias latipes)	9	NP_001265759.1	6973668-6975395	ENSORLG00000003809	3	RNF165B, LOXHD1B, PTGER4A, MRPS27, ZNF366, *, FABP1B.1, SMYD1B, LZST3B, UBOX5, AVP,
Nile Tilapia (Oreochromis niloticus)	LG12	NC_031977.1	26377097-26379172	ENSONIG00000015235	4	AVP, UBOX5, LZST3B, SMYD1B, FABP1B.1, *, ZNF366, MRPS27, PTGER4A, LOXHD1B, RNF165B
Zebrafish (Danio rerio)	5	NC_007116.7	72087619-72093255	ENSDARG00000042845	4	NUP214, FAM78AB, PLPP7,PRRC2B,DQX1,*,THNSL2,FABP1A,SMYD1A, SPRA, ABL1
Coelacanth (Latimeria chalumnae)	Unk	NW_005820294.1	484041-492783	ENSLACG00000020667	4	MAVS,PTPRA,GNRH2, *, AVP,UBOX5,FASTKD5, LZTS3
Elephant Shark (Callorhinchus milii)	Unk	NW_006890112.1	2519993-2523644	103181497	3	DDRGK1, LOC, FASTKD5, UBOX5, AVP, *, PTPRA, MAVS, PANK2, RNF24, SMOX
Japanese Lamprey (Lethenteron camtschaticum)				NOT PRESENT		
Sea Lamprey (Petromyzon marinus)				NOT PRESENT		

Table S1b: Microsynteny analysis (10-gene window) for OT.

Organism	Chromosome/ Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Adjacent genes
Human (Homo Sapiens)	3	NC_000003.12	8748579-8769614	ENSG00000180914	4	EDEM1,GRM7,LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Chimpanzee (Pan troglodytes)	3	NC_006490.4	8984327-9006633	ENSPTRG00000014582	4	EDEM1,GRM7,LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Western Gorilla (Gorilla gorilla)	3	NC_018427.2	8994821-9013929	ENSGGOG00000011995	4	EDEM1,GRM7,LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Northern white-cheeked Gibbon (Nomascus leucogenys)	21	NC_019836.1	59185191-59202387	ENSNLEG00000029699	5	EDEM1,GRM7,LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Rhesus Macaque (Macaca mulatta)	2	NC_027894.1	57649844-57664987	ENSMMUG00000009703	2	LHFPL4,SETD5,THUMPD3,SRGAP3,RAD18,*,CAV3,SSUH2,LMCD1,GRM7,EDEM1
Marmoset (Callithrix jacchus)	15	NC_013910.1	60672302-60687374	ENSCJAG00000015914	2	ARL8B,EDEM1,GRM7,LMCD1,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Mouse lemur (Microcebus murinus)	26	NC_033685.1	7175181-7186091	ENSMICG0000003813	4	EIF4E1B,TSPAN17,GC,LMCD1,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Mouse (Mus musculus)	6	NC_000072.6	112473684-112491308	ENSMUSG00000049112	4	EDEM1,GRM7,LMCD1,SSU2,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Prarie Vole (Microtus ochrogaster)	JH996431.1	NW_004949099.1	26344441-26356922	ENSMOCG00000002867	2	EDEM1,GRM7,LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3, SETD5,LHFPL4
Cow (Bos taurus)	22	AC_000179.1	17817626-17827292	ENSBTAG00000019772	2	LHFPL4,SETD5,THUMPD3,SRGAP3,RAD18,*,CAV3,SSUH2,LMCD1,GRM7,EDEM1
Yangtze River Dolphin (Lipotes vexillifer)	248	NW_006790426.1	391774-406885	ENSTTRG00000006041.1	2	LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3,SETD5
Horse (Equus caballus)	16	NC_009159.2	7978752-7995514	ENSECAG00000017844	2	LHFPL4,SETD5,THUMPD3,SRGAP3,RAD18,*,CAV3,SSUH2,LMCD1,GRM7,EDEM1
Dog (Canis lupus familiaris)	20	NC_006602.3	9358465-9380314	ENSCAFG00000005553	5	LHFPL4,SETD5,THUMPD3,SRGAP3,RAD18,*,CAV3,SSUH2,LMCD1,GRM7,EDEM1
Megabat (Pteropus vampyrus)	3394	NW_011888952.1	60991-76592	ENSPVAG00000012219.1	2	LHFPL4,SETD5,THUMPD3,SRGAP3,RAD18,*,CAV3,SSUH2,LMCD1
Platypus (Ornithorhynchus anatinus)	3706	not available	10369-12457	ENSOANG00000007720	2	SRGAP3,RAD18,*,CAV3
Chicken (Gallus gallus)	12	NC_006099.4	19319352-19322402	ENSGALG00000003138	2	ARL8B,EDEM1,GRM7,LMCD1,CAV3,*,RAD18,SRGAP3,THUMPD3,VHL,IRAK2
Anna's hummingbird (Calypte anna)	Unk	NW_007621624.1	1897466-1903129	103535338	3	ARL8B,EDEM1,GRM7,LMCD1,CAV3,*,RAD18,SRGAP3,THUMPD3,VHL,SEC13
Zebra finch (Taeniopygia guttata)	12	NC_011476.1	20974715-20977711	ENSTGUG00000010403	2	ARL8B,EDEM1,GRM7,LMCD1,CAV3,*,RAD18,SRGAP3,THUMPD3,VHL,IRAK2
American Alligator (Alligator mississippiensis)	Unk	NW_017713410.1	5560332-5576875	102561033	3	IRAK2,VHL,THUMPD3,SRGAP3,RAD18,*,CAV3,LMCD1,GRM7,EDEM1,ARL8B
Carolina anole-lizard (Anolis carolinensis)	GL343273.1	NW_003338820.1	1755210-1783840	ENSACAG00000011241	4	IRAK2,VHL,THUMPD3,SRGAP3,RAD18,* (end of scaffold)
Painted turtle (Chrysemys picta bellii)	Unk	NW_007281343.1	7589746-7604073	101940340	3	IRAK2,VHL,THUMPD3,SRGAP3,RAD18,*,CAV3,SSUH2,LMCD1,GRM7,EDEM1
Tropical clawed frog (Xenopus tropicalis)	4	NC_030680.1	120901659-120941037	ENSXETT00000003039.3	2	IRAK2,VHL,THUMPD3,SRGAP3,RAD18,*,CAV3.1,SSUH2,LMCD1,GRM7,EDEM1
Three-spined stickleback (Gasterosteus aculeatus)	27	not available	2651053-2659788	ENSGACG00000000914	3	FANCD2,TEX264A,GRM2A,PARP3,CAV3,*,RAD18,SRGAP3,THUMPD3,CSE1L,PBRM1
Southern platyfish (Xiphophorus maculatus)	20	NW_005372220.1	629765-640289	ENSXMAG00000001524	2	CSE1L,KCNB1,THUMPD3,SRGAP3,RAD18,*,CAV3,PARP3,GRM2,TEX264,FANCD2
Spotted Gar (Lepisosteus oculatus)	LG5	NC_023183.1	26282331-26296485	ENSLOCG00000010646	3	PBRM1,SMIM4,STAB1,NISCH,RAD18,*,CAV3,PARP3,RRP9,GRM2B,TEX264A
Japanese medaka (Oryzias latipes)	5	NP_001243561.1	1143553-1150698	ENSORLG00000000719	2	PBRM1,CSE1L,THUMPD3,SRGAP3,RAD18,*,CAV3,PARP3,GRM2A,TEX264A,FANCD2
Nile Tilapia (Oreochromis niloticus)	LG5	NC_031970.1	16279887-16288393	ENSONIG00000018982	2	PBRM1,CSE1L,THUMPD3,SRGAP3,RAD18,*,CAV3,PARP3,GRM2A,TEX264,FANCD2
Zebrafish (Danio rerio)	6	NC_007117.7	41956254-41974803	ENSDARG00000033956	3	TWF2,CISH,HEMK1,SRGAP3,RAD18,*,CAV3,PARP3,GRM2A,TEX264A,FANCD2
Coelacanth (Latimeria chalumnae)	JH126579.1	NW_005819028.1	3120271-3133054	ENSLACG00000018362	2	ARL8B,EDEM1,GRM7,LMCD1,SSUH2,CAV3,*,RAD18,SRGAP3,THUMPD3,VHL,KIAA0895L
Elephant Shark (Callorhinchus milii)	Unk	NW_006890068.1	7859487-7868340	103176734	3	SEMA3H,ZMYND10,RASSF1,TUSC2,RAD18,*,CAV3,SSUH2,GRIP2,SEC13,CAND2
Japanese lamprey (Lethenteron japonicum)	KE993674	not available	2386-2392	not available		CACNA2D2/1,GRM2/3,SEMA3G/A/B/D/C/E,LMCD1,SSUH2,*,TMEM5,SRGAP3/2/1,C12orf66,THUMPD3, GRIP2
Sea Lamprey (Petromyzon marinus)	27	PMZ_0003232-RA	6293186-6301948	not available	2	VHL,GRIP1/2,THUMPD3,SRGAP2/3,TMEM5,*,SSUH2,PRICKLE3,LMCD1,TFE3,SEMA3AA/AB/G/D,TEX 264

Table S1c: Microsynteny analysis (10-gene window) for OTR.

Organism	Chromosome/ Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Adjacent genes
Human (Homo Sapiens)	12	NC_000012.12	63142759-63152810	ENSG00000166148	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,DPY19L2,TMEM5,SRGAP1,C12orf66,C12orf56,XPOT,TBK1
Chimpanzee (Pan troglodytes)	12	NC_0006479.4	26547336-26557930	ENSPTRG00000005167	2	TBK1,XPOT,C12Horf56,C12Horf66,SRGAP1,TMEM5,DPY19L2,*,PPM1H,MON2,USP15,FAM19A2,SLC16 A7
Western Gorilla (Gorilla gorilla)	12	NC_018436.2	22873386-22883461	ENSGGOG00000037345	2	TBK1,XPOT,C12Horf56,C12Horf66,SRGAP1,TMEM5,DPY19L2,*,PPM1H,MON2,USP15,FAM19A2,SLC16 A7
Northern white-cheeked Gibbon (Nomascus leucogenys)	11	NC_019826.1	49513207-49517221	ENSNLEG00000017088	2	TBK1,XPOT,C12Horf56,C12Horf66,SRGAP1,TMEM5,DPY19L2,*,PPM1H,MON2,USP15,FAM19A2,SLC16 A7
Rhesus Macaque (Macaca mulatta)	11	NC_027903.1	62120324-62126735	ENSMMUG00000000549	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,DPY19L2,TMEM5,SRGAP1,C11Horf66,C11Horf56,XPOT,TB K1
Marmoset (Callithrix jacchus)	9	NC_013904.1	52391174-52399758	ENSCJAG00000006111	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,DPY19L2,TMEM5,SRGAP1,C9H12orf66,C9H12orf56,XPOT,T BK1
Mouse lemur (Microcebus murinus)	7	NC_033666.1	73514267-73519446	ENSMICG00000038397	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,DPY19L2,TMEM5,SRGAP1,C7H12orf66,C7H12orf56,XPOT,T BK1
Mouse (Mus musculus)	10	NC_000076.6	122448499-122453453	ENSMUSG00000020123	2	RASSF3,TBK1,XPOT,SRGAP1,TMEM5,*,PPM1H,MON2,USP15,FAM19A2,SLC16A7
Prarie Vole (Microtus ochrogaster)	24	NC_022024.1	4257324-4262940	ENSMOCG00000008797		TBC1D30,GNS,RASSF3,TBK1,XPOT,*,PPM1H,MON2,USP15,FAM19A2,SLC16A7
Cow (Bos taurus)	5	AC_000162.1	50592290-50595770	ENSBTAG00000007175	2	TBK1,XPOT,C5H12orf56,C5H12orf66,SRGAP1,TMEM5,DPY19L2,*,PPM1H,MON2,USP15,FAM19A2,SLC 16A7
Yangtze River Dolphin (Lipotes vexillifer)	99620	NW_006775771.1	38348-41865	ENSTTRG00000001865.1	2	SRGAP1,TMEM5,DPY19L2,*,PPM1H,MON2
Horse (Equus caballus)	6	NC_009149.2	79470270-79474773	ENSECAG00000010418	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,DPY19L2,TMEM5,SRGAP1,C6H12orf66,C6H12orf56,XPOT,T BK1
Dog (Canis lupus familiaris)	10	NC_006592.3	6266217-6269662	ENSCAFG00000000339	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,DPY19L2,TMEM5,SRGAP1,CUNH12orf66,,XPOT,TBK1
Megabat (Pteropus vampyrus)	5495	NW_011888790.1	44926-48380	ENSPVAG00000003341.1	2	*,TMEM5,SRGAP1,XPOT,TBK1
Platypus (Ornithorhynchus anatinus)	11199	NW_001603029.1	3605-9536	ENSOANG00000002085.2	2	only gene on contig
Chicken (Gallus gallus)	1	NC_006088.4	33298676-33301273	ENSGALT00000073363.1	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,TMEM5,SRGAP1,C1H12orf66,XPOT,TBK1
Anna's hummingbird (Calypte anna)				LIKELY PRESENT		scaffolding gap
Zebra finch (Taeniopygia guttata)	1A	NC_011463.1	32624808-32628422	ENSTGUG00000006366	3	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,TMEM5,SRGAP1,C1AH12orf66,XPOT,TBK1
American Alligator (Alligator mississippiensis)	Unk	NW_017707830.1	29683015-29687391	102559961	2	SLC16A7,FAM19A2,USP15,MON2,PPM1H,*,TMEM5,SRGAP1,CUNH12orf66,XPOT,TBK1
Carolina anole-lizard (Anolis carolinensis)	5	NC_014780.1	50861538-50877718	ENSACAG00000000443	2	TBK1,XPOT,RPL18A,C5H12orf66,SRGAP1,TMEM5,*,PPM1H,MON2,USP15,FAM19A2,SLC16A7
Painted turtle (Chrysemys picta bellii)	Unk	NW_007281411.1	2994322-3008117	101953349	2	TWIST1,FAM19A2,USP15,MON2,PPM1H,*,TMEM5,SRGAP1,XPOT,TBK1,RASSF3
Tropical clawed frog (Xenopus tropicalis)	3	NC_030679.1	42380832-42386602	ENSXETT00000040693.2	2	CDK17,C12orf63,NEDD1,TMPO,SLC25A3,NUAK1,*,PPM1H,MON2, RPS16,OTOGL,PTPRQ
Three-spined stickleback (Gasterosteus aculeatus)	groupXIX	not available	4554417-4557763	ENSGACG00000003589	2	IFITM5,PTDSS2, TMEM168B, BMT2, *, PPM1H, MON2,FBLN1,WNT7BB,PPARAB
Southern platyfish (Xiphophorus maculatus)	Unk	NW_005372536.1	35299-39621	ENSXMAG00000017820.1	2	UPF2,RAB3IP,TMEM19,SRGAP1,*,
Spotted Gar (Lepisosteus oculatus)	LG8	NC_023186.1	47458997-47465626	ENSLOCG00000016982	2	FAM180A,MTPN,CLG8H12orf66,SRGAP1,TMEM5,*,PPM1H,MON2,USP15,KDM5A,RAD52
Japanese medaka (Oryzias latipes)	6	not available	5755264-5757013	ENSORLG00000002126		SFTPD,BMT2,,TMEM168B,PTDSS2,CDKN1CB,*,PPM1H,MON2,FBLN1,WNT7BB,PPARAB
Nile Tilapia (Oreochromis niloticus)	LG7	NC_031972.1	52560352-52565206	ENSONIG00000008870	3	PPARAB,WNT7B,AKR1D1,MON2,PPM1H,*,BMT2,TMEM168B,PTDSS2,CDKN1CB,IFITM5
Zebrafish (Danio rerio)	25	NC_007136.7	1550220-1558366	ENSDARG00000077083	2	FAM96A,CALML4B,CLN6B,FEM1B,ITGA11B,*,PPM1H,MON2,FBLN1,WNT7BB,PPARAB
Coelacanth (Latimeria chalumnae)	JH127196.1	NW_005819645.1	1049217-1052699	ENSLACG00000014523	2	FAM19A2,USP15,MON2,PPM1H,*
Elephant Shark (Callorhinchus milii)	39	NW_006890092.1	516987-525581	SINCAMT00000020158	2	FAM19A2,USP15,MON2,PPM1H,*,TMEM5,SRGAP1,C12orf66,C12orf56,XPOT,TBK1
Japanese lamprey (Lethenteron japonicum)	KE993677	not available	2802-2815	not available		GRIP1/2,LTA4H,TCAF2,CDK17/16/18,DNAH3/7/12/1,SRGAP1/2/3,*, KIAA1033, ALDH1L2,CNTN5/2,NFASC
Sea Lamprey (Petromyzon marinus)	10	PMZ_0013447-RA	12654136..12675190	not available	5	NFASC,NRCAM,CNTN2,ALDH1L2,Kiaa1033,*,SRGAP1,DNAH3/1/7,CAMTA2,TTC25,ABT1

Table S1d: Microsynteny analysis (10-gene window) for VTR1A.

Organism	Chromosome/Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Ajacent genes
Human (Homo Sapiens)	1	NC_000001.11	206109849-206117048	ENSG00000198049	2	PM20D1,SLC26A9,RAB7B,CTSE,C1orf186,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Chimpanzee (Pan troglodytes)	1	NC_006468.4	184998159-185008779	ENSPTRG00000023708	2	PM20D1,SLC26A9,RAB7B,CTSE,C1Horf186,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Western Gorilla (Gorilla gorilla)	1	NC_018424.2	185991853-185999451	ENSGGOG00000013058	2	RAB29,SLC41A1,PM20D1,SLC26A9,FAM72A,*,C1Horf186,CTSE,RAB7B,SRGAP2,IKBKE
Gibbon (Nomascus leucogenys)	5	NC_019820.1	55640176-55647901	ENSNLEG00000000448	2	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,C5H1orf186,CTSE,SLC26A9,PM20D1,SLC41A1
Rhesus Macaque (Macaca mulatta)	1	NC_027893.1	160482394-160488726	ENSMMUG00000002255	2	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,C1H1orf186,CTSE,SLC26A9,PM20D1,SLC41A1
Marmoset (Callithrix jacchus)	19	NC_013914.1	28140605-28150207	ENSCJAG00000047274b	3	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,C19H1orf186,CTSE,SLC26A9,PM20D1,SLC41A1
Mouse lemur (Microcebus murinus)	27	NC_033686.1	29843828-29850128	ENSMICG00000047529	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,C27H1orf186,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Mouse (Mus musculus)	1	NC_000067.6	131599114-131612000	ENSMUSG00000026432	2	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,CTSE,RAB7B,SLC26A9,PM20D1,SLC41A1
Prarie Vole (Microtus ochrogaster)	6	NC_022013.1	22094859-22103639	ENSMOCG00000002679	2	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,CTSE,SLC26A9,PM20D1,SLC41A1,RAB29
Cow (Bos taurus)	16	AC_000173.1	3726354-3731874	ENSBTAG00000017301	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE, C16H1orf186,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Yangtze River Dolphin (Lipotes vexillifer)	110735	NW_006798649.1	6994-79071	ENSTTRG00000009361	2	RAB7L1,SLC41A1,PM20D1,SLC26A9,RAB7B,LOC,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Horse (Equus caballus)	5	NC_009148.2	2416640-2424080	ENSECAG00000013720	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,C5H1orf186,*,SRGAP2,IKBKE,RASSF5,EIF2D,DYRK3
Dog (Canis lupus familiaris)	38	NC_006620.3	2440615-2446037	ENSCAFG00000010214	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,C38H1orf186,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Megabat (Pteropus vampyrus)	3164	NW_011888998.1	7745-9572	ENSPVAG00000000724.1	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,LOC,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Platypus (Ornithorhynchus anatinus)	Unk	NW_001693435.1	3819-8220	100080825	2	CNTN2,TMEM81,TMCC2,LOC,LOC,LOC,SLC26A9,RAB7B, *, SRGAP2,IKBKE,IL10,YOD1,SNORA7
Chicken (Gallus gallus)	26	NC_006113.4	2366153-2368135	ENSGALG00000000788	2	SLC41A1,NUCKS1,PM20D1,RAB7B,CTSE,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Anna's hummingbird (Calypte anna)	Unk	NW_007621465.1	318962-320774	103534715	3	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,CTSE,RAB7B,SLC26A9,PM20D1,SLC41A1
Zebra finch (Taeniopygia guttata)	26	NW_002198144.1	89404-92009	ENSTGUG00000017506	2	SLC41A1,NUCKS1,PM20D1,RAB7B,CTSE,*
American Alligator (Alligator mississippiensis)	Unk	NW_017713700.1	695343-700446	102568245	2	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,CUNH1orf186,CTSE,RAB7B,SLC26A9,PM20D1,SLC41A1
Carolina anole-lizard (Anolis carolinensis)	Unk	NW_003338775.1	988987-997249	100562094	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Painted turtle (Chrysemys picta bellii)	Unk	NW_007281457.1	1690896-1696595	101941797	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Tropical clawed frog (Xenopus tropicalis)	2	NC_030678.1	65411505-65415512	100487856	2	SLC41A1,PM20D1,SLC26A9,RAB7B,CTSE,*,FAM72A,SRGAP2,IKBKE,RASSF5,EIF2D
Three-spined stickleback (Gasterosteus aculeatus)	27			DELETED		RASSF5,IKBKE,SRGAP2,FAM72B,FOXP4,MDF1,TFEB, TMEM183A
Southern platyfish (Xiphophorus maculatus)	20			DELETED		RASSF5,IKBKE,SRGAP2,FAM72B,FOXP4,MDF1,TFEB, TMEM183A
Spotted Gar (Lepisosteus oculatus)	LG3			DELETED		RASSF5,IKBKE,SRGAP2,FAM72A,FOXP4,MDF1,TFEB, TMEM183A
Japanese medaka (Oryzias latipes)	5			DELETED		RASSF5,IKBKE,SRGAP2,FAM72A,FOXP4,MDF1,TFEB, TMEM183A
Nile Tilapia (Oreochromis niloticus)	GL831146			DELETED		RASSF5,IKBKE,SRGAP2,FAM72B,FOXP4,MDF1,TFEB, TMEM183A
Zebrafish (Danio rerio)	11			DELETED		RASSF5,IKBKE,SRGAP2,FAM72B,FOXP4,MDF1,TFEB, TMEM183A
Coelacanth (Latimeria chalumnae)	JH127167.1	NW_005819616.1	895053-901723	ENSLACG00000013680	2	EIF2D,RASSF5,IKBKE,SRGAP2,FAM72A,*,CTSE,RAB7B,LOC,PM20D1,SLC41A1
Elephant Shark (Callorhinchus milii)	Unk	NW_006890123.1	417088-419796	SINCAMG00000008883	2	ATXN7L2,SYPL2B,TBK1,SRGAP2,FAM72B,*,TFEB,MDF1,FOXP4,KCNC4,PROK1,LAMTOR5
Japanese Lamprey (Lethenteron						

Table S1e: Microsynteny analysis (10-gene window) for VTR1B.

Organism	Chromosome/ Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Adjacent genes
Human (Homo Sapiens)	X	NC_000023.11	153902531-153907166	ENSG00000126895	5	IDH3G,SSR4,PDZD4,L1CAM,LCA10,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Chimpanzee (Pan troglodytes)	X	NC_006491.4	153571841-153575085	ENSPTRG00000033849	4	IDH3G,SSR4,PDZD4,L1CAM,LCA10,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Western Gorilla (Gorilla gorilla)	X	NC_018447.2	154243887-154246414	ENSGGOG00000027387	4	PLXNB3,SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Gibbon (Nomascus leucogenys)	X	NC_019841.1	139620495-139622932	ENSNLEG00000013671	4	PLXNB3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Rhesus Macaque (Macaca mulatta)	X	NC_027913.1	147380143-147382642	ENSMMUG00000045962	4	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Marmoset (Callithrix jacchus)	X	NC_013918.1	140442283-140443346	ENSCJAG00000011022	2	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Mouse lemur (Microcebus murinus)	X	NC_033692.1	74918489-74922240	ENSMICG00000037091	3	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Mouse (Mus musculus)	X	NC_000086.7	73891798-73894428	ENSMUSG00000031390	4	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,IRAK1
Prarie Vole (Microtus ochrogaster)	X	NW_004949227.1	74136-76404	ENSMOCG00000012257	3	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1
Cow (Bos taurus)	X	AC_000187.1	40035987-40038825	ENSBTAG00000047138	4	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Yangtze River Dolphin (Lipotes vexillifer)	3422	NW_006784311.1	81275-82922	ENSTTRG00000007453	3	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Horse (Equus caballus)	X	NC_009175.2	122355333-122357852	ENSECAG00000022974	4	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Dog (Canis lupus familiaris)	X	NC_006621.3	121752283-121754794	NSCAFG00000019402	5	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Megabat (Pteropus vampyrus)	3867	NW_011889056.1	45661-47320	ENSPVAG00000000461.1	4	SRPK3,IDH3G,SSR4,PDZD4,L1CAM,*,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
Platypus (Ornithorhynchus anatinus)	8921	NW_001786488.1	10954-15088	100091389	4	CAV2,* (?)
Chicken (Gallus gallus)				DELETED		both receptor and territory are deleted
Anna's hummingbird (Calypte anna)				DELETED		both receptor and territory are deleted
Zebra finch (Taeniopygia guttata)				DELETED		both receptor and territory are deleted
Carolina anole-lizard (Anolis carolinensis)	2	NC_014777.1	88181923-88196672	ENSACAG00000021016	3	HAUS7,FLNA,SSR4,PDZD4,L1CAM,*,CAV3,ARHGAP4,NAA10,RENBP,HCFC1,TMEM187
American Alligator (Alligator mississippiensis)				LIKELY PRESENT		scaffolding gap
Painted turtle (Chrysemys picta bellii)	Unk	NW_007284486.1	166-940	101946394	2	LOC,LOC,*
Tropical clawed frog (Xenopus tropicalis)	8	NC_030684.1	12407743-12454885	100487051	4	SLC31A2,TMEM187,IRAK1,MECP2,OPN1LW,*,BRINP1,ASTN2,TRIM32,PAPPA,ARHGAP4
Three-spined stickleback (Gasterosteus aculeatus)	groupXII	unk	10794244-10798372	ENSGACG00000008681	7	ADCY6B,GALNT6,NAA10,ARHGAP4A,LOC,SSR4, *,IDH3G,FAM3A,WNK3,PHF8,HUWE1
Southern platyfish (Xiphophorus maculatus)	JH556723.1:	NW_005372241.1	1303486-1322484	ENSXMAG00000010783	6	ADCY6B,GALNT6,LOC,ARHGAP4A,LOC,SSR4, *,IDH3G,FAM3A,WNK3,PHF8,HUWE1
Spotted Gar (Lepisosteus oculatus)	AHAT010435 08.1	NW_006270651.1	232-3820	ENSLOCG00000000338	3	Only gene on contig
Japanese medaka (Oryzias latipes)	7	not available	11847350-11854364	ENSORLG00000007088	7	CACNB3B,ADCY6B,NAA10,SSR4, *,IDH3G,FAM3A,WNK3,PHF8,HUWE1
Nile Tilapia (Oreochromis niloticus)	LG20	NC_031984.1	24721380-24735737	ENSONIG00000012001	8	ADCY6B,GALNT6,NAA10,ARHGAP4A,LOC,SSR4, *,IDH3G,FAM3A,WNK3,PHF8,HUWE1
Zebrafish (Danio rerio)	23	NC_007134.7	25037773-25067138	ENSDARG00000007436	6	KLHL21,ZBTB48,NOL9,CASP9, ARHGAP4A,*,IDH3G,FAM3A,ERBB3B,PA2G4B
Coelacanth (Latimeria chalumnae)	Unk	NW_005819525.1	1126070-1202567	ENSLACG00000014898	2	BCAP31,SLC6A8,PDZD4,L1CAM,*,
Elephant Shark (Callorhinchus milii)				NOT PRESENT		
Japanese Lamprey (Lethenteron camtschaticum)				NOT PRESENT		

Table S1f: Microsynteny analysis (10-gene window) for VTR2A.

Organism	Chromosome/Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Ajacent genes
Human (Homo Sapiens)				DELETED		territory is present but the receptor deleted
Chimpanzee (Pan troglodytes)				DELETED		territory is present but the receptor deleted
Western Gorilla (Gorilla gorilla)				DELETED		territory is present but the receptor deleted
Gibbon (Nomascus leucogenys)				DELETED		territory is present but the receptor deleted
Rhesus Macaque (Macaca mulatta)				DELETED		territory is present but the receptor deleted
Marmoset (Callithrix jacchus)				DELETED		territory is present but the receptor deleted
Mouse lemur (Microcebus murinus)				DELETED		territory is present but the receptor deleted
Mouse (Mus musculus)				DELETED		territory is present but the receptor deleted
Prarie Vole (Microtus ochrogaster)				DELETED		territory is present but the receptor deleted
Cow (Bos taurus)				DELETED		territory is present but the receptor deleted
Yangtze River Dolphin (Lipotes vexillifer)				DELETED		territory is present but the receptor deleted
Horse (Equus caballus)				DELETED		territory is present but the receptor deleted
Dog (Canis lupus familiaris)				DELETED		territory is present but the receptor deleted
Megabat (Pteropus vampyrus)				DELETED		territory is present but the receptor deleted
Platypus (Ornithorhynchus anatinus)				DELETED		territory is present but the receptor deleted
Chicken (Gallas gallas)				DELETED		territory is present but the receptor deleted
Anna's hummingbird (Calypte anna)				DELETED		territory is present but the receptor deleted
Zebra finch (Taeniopygia guttata)				DELETED		territory is present but the receptor deleted
American Alligator (Alligator mississippiensis)				DELETED		territory is present but the receptor deleted
Carolina anole-lizard (Anolis carolinensis)				DELETED		territory is present but the receptor deleted
Painted turtle (Chrysemys picta bellii)				DELETED		territory is present but the receptor deleted
Tropical clawed frog (Xenopus tropicalis)				DELETED		territory is present but the receptor deleted
Three-spined stickleback (Gasterosteus aculeatus)	scaffold_27	not available	2069125-2070759	ENSGACG00000000754	2	RPS26, FKBP11, ARF3B, ERBB3A, CRBN, *, VHL, TATDN2, CCDC174, FGD5B, PRKCDA
Southern platyfish (Xiphophorus maculatus)	JH556702.1	not available	1472609-1477351	ENSXMAG00000001906	3	RTF1, PRKCDA, CCDC174, TATDN2, VHL, *, CRBN, PAG2G4A, ERBB3, ARF3B, FKBP11, RPS26
Spotted Gar (Lepisosteus oculatus)	LG5	NC_023183.1	28118103-28120790	ENSLOCG00000011096	3	SUMF1, LOC, LRRN1, SNX6, DNAJB9L, CRBN, *, TRNT1, VHL, TATDN2, GHRL, CCDC174
Japanese medaka (Oryzias latipes)	5	NM_001278816	2182673-2184327	ENSORLG00000001494	3	RTF1, PRKCD, CCDC174, TATDN2, VHL, *, CRBN, ERBB3, FKBP11, RPS26, IKZF4
Nile Tilapia (Oreochromis niloticus)	GL831146.1	not available	2358757-2360693	ENSONIG00000019049	2	RTF1, FGD5B, CCDC174, TATDN2, VHL, *, CRBN, PAG2G4A, ERBB3, FKBP11, RPS26, IKZF4
Zebrafish (Danio rerio)	6		40,4-40,6	-		CCDC174, TATDN2, GHRL, VHL, CRBN,PRKCDA
Coelacanth (Latimeria chalumnae)	JH127297.1	NW_005819746.1	863612-866135	ENSLACG00000013474	2	LOC, *, TRNT1, CRBN, LOC
Elephant Shark (Callorhinchus milii)	KI635990	NW_006890189.1	1733811-1744399	KI635990.52	5	EDEM1, ARL8BA, ITPR1, SUMF1, LRRN1, SNX6, *, CRBN, LOC, TRNT1, CHL1, CNTN4
Japanese Lamprey (Lethenteron camtschaticum)	KE993674	not available	6843-6850	not available		KBTBD8, SLC16A7, RHOA, EMC3, SORT1, *, CRBN, SLC25A26, DMTF1, MANF, FRMD4B/A
Sea Lamprey (Petromyzon marinus)	27	PMZ_0008155-RA	640528-647515	not available	5	SLC16A1, RHOA, EMC3, GPX2, SORCS1, *, CRBN, DMTF1, MANF, FAM107B, FRMD4B/A

Table S1g: Microsynteny analysis (10-gene window) for VTR2B.

Organism	Chromosome/Scaffold	NCBI Accession	Locus	Ensembl ID/GeneID	# of Exons	Adjacent genes
Human (Homo Sapiens)	7			DELETED		territory is present but the receptor deleted
Chimpanzee (Pan troglodytes)	7			DELETED		territory is present but the receptor deleted
Western Gorilla (Gorilla gorilla)	7			DELETED		territory is present but the receptor deleted
Gibbon (Nomascus leucogenys)	13			DELETED		territory is present but the receptor deleted
Rhesus Macaque (Macaca mulatta)	3			DELETED		territory is present but the receptor deleted
Marmoset (Callithrix jacchus)	8			DELETED		territory is present but the receptor deleted
Mouse lemur (Microcebus murinus)	11			DELETED		territory is present but the receptor deleted
Mouse (Mus musculus)	12			DELETED		territory is present but the receptor deleted
Prarie Vole (Microtus ochrogaster)	1			DELETED		territory is present but the receptor deleted
Cow (Bos taurus)	4			DELETED		territory is present but the receptor deleted
Yangtze River Dolphin (Lipotes vexillifer)	Unk			DELETED		territory is present but the receptor deleted
Horse (Equus caballus)	4			DELETED		territory is present but the receptor deleted
Dog (Canis lupus familiaris)	18			DELETED		territory is present but the receptor deleted
Megabat (Pteropus vampyrus)	1606			DELETED		territory is present but the receptor deleted
Platypus (Ornithorhynchus anatinus)	Unk			DELETED		territory is present but the receptor deleted
Chicken (Gallus gallus)	1	NC_006088.4	28501163-28508107	ENSGALG00000009497	3	DOCK4, LRRN3, IMMP2L, DNAJB9, THAP5,*, PNPLA8, NRCAM, CNTN1, PDZRN4, GXYLT1
Anna's hummingbird (Calypte anna)	Unk	NW_007619672.1	15544-17868	not available	2	*, PNPLA8, NRCAM, CNTN1 (only genes on scaffold)
Zebra finch (Taeniopygia guttata)	1A	NC_011463.1	27785040-27787243	ENSTGUG00000005614	2	DOCK4, IMMP2L, LRRN3, DNAJB9, THAP5,*, PNPLA8, NRCAM, CNTN1, PDZRN4, GXYLT1
American Alligator (Alligator mississippiensis)	Unk	NW_017707830.1	20671392-20677406	102570746	2	DOCK4, LOC, LOC, IMMP2L, LOC, DNAJB9, THAP5,*, LOC, PNPLA8, NRCAM, CNTN1, PDZRN4, GXYLT1
Green anole-lizard (Anolis carolinensis)	5	NC_014780.1	57154865-57166337	ENSACAG00000025984	2	GXYLT1, PDZRN4, CNTN1, NRCAM, PNPLA8, *LOC103278908, LOC, LOC, IMMP2L, LRRN3, DOCK4
Painted turtle (Chrysemys picta bellii)	Unk	NW_007359870.1	7082334-7086401	101945660	2	GXYLT1, LOC, PDZRN4,CNTN1,NRCAM,PNPLA8,*,THAP5,DNAJB9,IMMPL2,LRRN3, DOCK4
Three-spined stickleback (Gasterosteus aculeatus)	IV	not available	18799914-18802428	ENSGACG00000018884	3	PDZRN4, SNORA74, CNTN1B, NRCAMA, PNLA8, *, LOC, DNAJB9B, AKR1B1, LOC, TFEC, FOXP2, GPR85
Tropical clawed frog (Xenopus tropicalis)	3	NC_030679.1	55974203-55981894	ENSXETG00000015741	2	GXYLT1, PDZRN4,CNTN1,NRCAM,PNPLA8,*,THAP5,DNAJB9,IMMPL2,LRRN3, LOC, LOC, DOCK4
Southern platyfish (Xiphophorus maculatus)	JH556693.1	NW_005372211.1	1843877-1845656	ENSXMAG00000009218	2	FOXP2, MDFIC, TFEC, AKR1B1, DNAJB9B, *, PNPLA8, NRCAMA, CNTN1B, PDZRN4, GXYLT1B
Spotted Gar (Lepisosteus oculatus)	Unk	NW_006270074.1	467850-476904	ENSLOCG00000001076	2	DNAJB9, THAP5,*, PNPLA8, NRCAM, CNTN1, LOC, PDZRN4, GXYLT1
Japanese medaka (Oryzias latipes)	23			DELETED		CNTN1B,NRCAMA,PNPLA8,LOC,DNAJB9B,LOC,AKR1B1
Nile Tilapia (Oreochromis niloticus)	GL831226.1			DELETED		TFEC,LOC,LOC,DNAJB9B,THAP5,PNPLA8,NRCAMA
Zebrafish (Danio rerio)	4			DELETED		PNPLA8, NRCAMA, CNTN1B, PDZRN4, GXYLT1B
Coelacanth (Latimeria chalumnae)	JH127820			DELETED		NRCAM,PNPLA8,THAP5,DNAJB9
Elephant Shark (Callorhinchus milii)	scaffold_136	NW_006890189.1	1733245-1735049	SINCAMG00000000562	1	DOCK4, IMMPL2, LRRN3, DNAJB9, THAP5, *, PNPLA8, SLC6A1, EFCAB6, NAMPT
Japanese Lamprey (Lethenteron camtschaticum)	KE993677	not available	3496-3514	not available		KIAA1033,ALDH1L2,CNTN2/5/3/4/1/6,NFASC,*,IMMP2L,LRRN1/3/2,PRKCD,FLNA/B/C,SLMAP
Sea Lamprey (Petromyzon marinus)	scaf_00010	PMZ_0042163-RA	11954340-11969247	not available	2	SRGAP,VTR1A,KIAA1033,CNTN2,NFASC/NRCAM,*,IMMP2L,LRRN1/3/2,DENND6A,GCC1,SLMAP.

Table S1h: Microsynteny analysis (10-gene window) for VTR2C.

Organism	Orientation
Human (Homo Sapiens)	tail-to-tail
Chimpanzee (Pan troglodytes)	tail-to-tail
Western Gorilla (Gorilla gorilla)	tail-to-tail
Gibbon (Nomascus leucogenys)	tail-to-tail
Rhesus Macaque (Macaca mulatta)	tail-to-tail
Marmoset (Callithrix jacchus)	tail-to-tail
Mouse lemur (Microcebus murinus)	tail-to-tail
Mouse (Mus musculus)	tail-to-tail
Prarie Vole (Microtus ochrogaster)	tail-to-tail
Cow (Bos taurus)	tail-to-tail
Yangtze River Dolphin (Lipotes vexillifer)	tail-to-tail
Horse (Equus caballus)	tail-to-tail
Dog (Canis lupus familiaris)	tail-to-tail
Megabat (Pteropus vampyrus)	tail-to-tail
Platypus (Ornithorhynchus anatinus)	tail-to-tail
Kangaroo rat (Dipodomys ordii)	tail-to-tail
Chicken (Gallus gallus)	tail-to-head
Anna's hummingbird (Calypte anna)	tail-to-head
Zebra finch (Taeniopygia guttata)	tail-to-head
Green anole-lizard (Anolis carolinensis)	tail-to-head
American Alligator (Alligator mississippiensis)	tail-to-head
Painted turtle (Chrysemys picta bellii)	tail-to-head
Three-spined stickleback (Gasterosteus aculeatus)	tail-to-head
Tropical clawed frog (Xenopus tropicalis)	tail-to-head
Southern platyfish (Xiphophorus maculatus)	tail-to-head
Spotted Gar (Lepisosteus oculatus)	tail-to-tail
Japanese medaka (Oryzias latipes)	tail-to-head
Nile Tilapia (Oreochromis niloticus)	tail-to-head
Zebrafish (Danio rerio)	tail-to-head
Coelacanth (Latimeria chalumnae)	tail-to-head
Elephant Shark (Callorhinchus milii)	tail-to-head

Table S2: Orientation of OT and VT in the species included in this study. Tail-to-head means that the OT has the same orientation as VT; tail-to-tail means that the OT is inverted relative to VT.

Gene	Chromosome	NCBI Accession	Locus	Ensembl ID	# of Exons	Adjacent genes
OTRa	6	NC_007117.7	41956254-41974803	ENSDARG00000033956	3	TWF2,CISH,HEMK1,SRGAP3,RAD18,OTRa,CAV3,PARP3,GRM2A,TEX264A,FANCD2
OTRb	6	NM_001199369	46668717-46687079	ENSDARG00000044175	2	PBRM1L,STAU1,RDH20,LOC,KCNB1, PTGIS,OTRb,TESPA1,TARBP2,ZGC,IGFN1.2,IGFN1,4
VTR1Aa	25	NC_007136.7	1550220-1558366	ENSDARG00000077083	2	FAM96A,CALML4B,CLN6B,FEM1B,ITGA11B,V1Aa,PPM1H,MON2,FBLN1,WNT7BB,PPARAB
VTR1Ab	4	NM_001297676	9207591-9214419	ENSDARG00000045788	2	SAMM50L,ALDH1L2,LOC,CHST11,NFYBA,HCFC2,V1Ab,SRGAP1B,LOC,IPO8,TMTC1,IMF2B,KTLGB
VTR1B	11			DELETED		RASSF5,IKBKE,SRGAP2,FAM72B,FOXP4,MDF1,TFEB, TMEM183A
VTR2Aa	23	NC_007134.7	25037773-25067138	ENSDARG00000007436	6	KLHL21,ZBTB48,NOL9,CASP9, ARHGAP4A,V2Aa,IDH3G,FAM3A,ERBB3B,PA2G4B
VTR2Ab	23	NC_007134.7	18563445-18580428	ENSDARG00000029219	8	NUCKS1B,FAM19A4A,FAM19A1A,HSD17B10,FAM120C,V2Ab,SEPHS2,SSR4,IKBKG,LOC,ARHGAP4B
VTR2Ba	6		40.4-40.6			CCDC174, TATDN2, GHRL, VHL, CRBN,PRKCDA
VTR2Bb	6	NC_007117.7	43218636-43221983	ENSDARG00000076797	2	ARL8BA,TNFRSF18,TCTA,GLYCTK,LRRN1,LOC,V2Bb,TRNT1,ARL6IP5A,FRMD4BA,LOC,MITFA,EEVS
VTR2Ca	4		13.6-13.8			PNPLA8, NRCAMA, CNTN1B, PDZRN4, GXYLT1B
VTR2Cb	4	NC_007115.7	17835557-17838335	ENSDARG00000076690	2	GNPTAB, CHPT1, MYBPC1, LOC, SPI2, SPIC, DNAJB9B, *, APAF1, ANKS1B, UHRF1BP1L, PLXNC1, CRADD

Table S3: All OTR-VTR present in zebrafish. First gene copies, namely the ones shared with the rest of the vertebrates, are indicated with ‘a’ (e.g. VTR1Aa). Second gene copies, namely the ones that were formed after the teleost-specific WGD, are indicated with ‘b’ (e.g. VTR1Ab).

Simrbase-ID	Putative gene in vertebrates	Super scaffold	Locus	# of Exons	Gene length	Adjacent genes
PMZ_0045207-RA	-	49	3426470-3434237	2	7768	NMRK1,LMOD3,FOXP4,MITF,GATA2,*,YAF2,PDZRN4,CLPS,CNTN2/3,NADL1.1.
PMZ_0032217-RA		49	3434273-3437054	2	2782	
				4	10584	
PMZ_0014716-RA	-	49	4609394-4620022	4	10629	NFASC,PNPLA8,WNT7B,FBLN1,WNK1,*,TRNT1,LRRN1,FGD3,SLC23A1,DUSP7
PMZ_0003232-RA	OTR	27	6293186-6301948	2	8763	VHL,GRIP1/2,THUMPD3,SRGAP2/3,TMEM5,*,SSUH2,PRICKLE3,LMCD1,TFE3,SEMA3AA/AB/G/D
PMZ_0008155-RA	VTR2B/V2B	27	640528-647515	5	6988	SLC16A1, RHOA, EMC3, GPX2, SORCS1, *, CRBN, DMTF1, MANF, FAM107B, FRMD4B/A
PMZ_0013447-RA	VTR1A/V1A	10	12654136-12675190	5	21055	NFASC,NRCAM,CNTN2,ALDH1L2,Kiaa1033,*,SRGAP1,DNAH3/1/7,CAMTA2,TTC25,ABT1
PMZ_0042163-RA	VTR2C/V2C	10	11954340-11969247	3	14908	KIAA1033,ALDH1L2,CNTN2,NFASC,*,IMMP2L,LRRN1,DENDD6A,GCC1,SLMAP

Simrbase-ID	Ensemble ID	Scaffold	Locus	# of Exons	Gene length	Adjacent genes
PMZ_0045207-RA	ENSPMAG00000009764	GL479461	3925-14587	5	10662	only gene on contig
PMZ_0032217-RA						
PMZ_0014716-RA	-	-	-	-	-	-
PMZ_0003232-RA	ENSPMAG00000001242	GL477135	197258-202768	3	5510	SSUH2rs1,*,FBXO25,?,?,?,THUMPD3
PMZ_0008155-RA	ENSPMAG00000002106	GL476972	181639-188819	4	7180	?,SORT1,*
PMZ_0013447-RA	ENSPMAG00000007650	GL478345	24256-33787	4	9531	only gene on contig
PMZ_0042163-RA	ENSPMAG00000003512	GL481097	22590-23695	4	1105	only gene on contig

Table S4: Top table shows all the putative OTR-VTRs in the sea lamprey germline genome and their putative orthology to OTR-VTR in the rest of the vertebrates. Bottom table shows a possible correspondence of the OTR-VTR found in the germline genome (Simrbase ID) to the ones found in the somatic genome (Ensembl ID).

Putative gene	Contig	Locus	Ensembl ID	# of Exons	Adjacent putative genes
VT	FYBX02010345.1	6598-6593	ENSEBUG00000007861	7	*, ?, rnf38, papola/papolg, esrra,f5,selp,wrap53
VTR	FYBX02010841.1	2971478-2997887	ENSEBUG00000007964	3	skube1/3,slc35c1,scyl2/trak2,tusc2,ier51,eif2d/rassf1,frmd4a,asb15b,*, kcnq3/2b
VTR	FYBX02010521.1	1075197-1157936	ENSEBUG00000001467	3	pdzrn4, gxylt2, *, itih4, slc6a6,grip2a, dstyk

Table S5: Putative VT ligand and VTR in inshore hagfish. Many genes of the 'Adjacent putative genes' were not annotated in the inshore hagfish genome. We used the 'Region comparison' tool of Ensembl to map each gene of these contigs against the human, zebrafish and sea lamprey genomes, and we have written here all the results of this search.

Human_OTR	Max Score	Identity	Total score	E-value
Chicken_OTR	682	78%	1280	0.0
Chicken_VTR1A	329	71%	781	4,00E-92
Chicken_VTR1B	210	67%	694	2,00E-56
Chicken_VTR2A				
Chicken_VTR2B				
Chicken_VTR2C				
Frog_OTR	354	71%	1131	1,00E-98
Frog_VTR1A	114	65%	378	3,00E-27
Frog_VTR1B	66.2	82%	94.5	1,00E-12
Frog_VTR2A				
Frog_VTR2B				
Frog_VTR2C	46.4	75%	205	2,00E-06
Spotted gar_OTR	246	67%	579	2,00E-66
Spotted gar_VTR1A	188	69%	477	2,00E-49
Spotted gar_VTR1B				
Spotted gar_VTR2A	107	75%	477	3,00E-25
Spotted gar_VTR2B	48.2	77%	341	2,00E-07
Spotted gar_VTR2C	42.8	80%	371	3,00E-05
Coelacanth_OTR	432	72%	882	1,00E-122
Coelacanth_VTR1A	59	76%	301	1,00E-10
Coelacanth_VTR1B	73.4	87%	310	1,00E-14
Coelacanth_VTR2A	91.5	70%	463	5,00E-19
Coelacanth_VTR2B	50	66%	170	5,00E-08
Coelacanth_VTR2C				
Elephant shark_OTR	324	69%	721	5,00E-90
Elephant shark_VTR1A	214	70%	613	6,00E-57
Elephant shark_VTR1B	230	68%	637	3,00E-62
Elephant shark_VTR2A				
Elephant shark_VTR2B	75.2	69%	397	1,00E-15
Elephant shark_VTR2C				
Sea lamprey_OTR	316	69%	510	7,00E-88
Sea lamprey_VTR1A	333	75%	686	2,00E-92
Sea lamprey_VTR1B				
Sea lamprey_VTR2A				
Sea lamprey_VTR2B	80.6	69%	298	8,00E-17
Sea lamprey_VTR2C	158	68%	436	8,00E-40

Human_VTR1A	Max Score	Identity	Total score	E-value
Chicken_OTR	179	65%	301	2,00E-47
Chicken_VTR1A	738	80%	1006	0.0
Chicken_VTR1B	282	69%	459	2,00E-78
Chicken_VTR2A				
Chicken_VTR2B				
Chicken_VTR2C	60.8	67%	602	3,00E-11
Frog_OTR	96.9	72%	434	3,00E-21
Frog_VTR1A	385	70%	725	6,00E-109
Frog_VTR1B	82.4	69%	494	6,00E-18
Frog_VTR2A	64.4	68%	648	2,00E-11
Frog_VTR2B				
Frog_VTR2C	41	77%	558	4,00E-05
Spotted gar_OTR	102	75%	571	2,00E-23
Spotted gar_VTR1A	360	70%	880	3,00E-101
Spotted gar_VTR1B				
Spotted gar_VTR2A	95.1	69%	278	9,00E-22
Spotted gar_VTR2B	69.8	68%	360	3,00E-14
Spotted gar_VTR2C				
Coelacanth_OTR	69.8	72%	539	1,00E-13
Coelacanth_VTR1A	174	65%	464	1,00E-45
Coelacanth_VTR1B	84.2	75%	252	3,00E-18
Coelacanth_VTR2A	77	70%	659	5,00E-15
Coelacanth_VTR2B	50	66%	145	2,00E-08
Coelacanth_VTR2C				
Elephant shark_OTR	104	68%	349	4,00E-24
Elephant shark_VTR1A	224	70%	497	2,00E-60
Elephant shark_VTR1B	235	68%	417	3,00E-64
Elephant shark_VTR2A				
Elephant shark_VTR2B	89.7	70%	263	3,00E-20
Elephant shark_VTR2C				
Sea lamprey_OTR	235	67%	412	1,00E-63
Sea lamprey_VTR1A	313	72%	609	1,00E-86
Sea lamprey_VTR1B				
Sea lamprey_VTR2A				
Sea lamprey_VTR2B	86	78%	422	9,00E-19
Sea lamprey_VTR2C	131	69%	374	5,00E-32

Human_VTR1B	Max Score	Identity	Total score	E-value
Chicken_OTR	250	67%	653	1,00E-68
Chicken_VTR1A	279	72%	1211	2,00E-77
Chicken_VTR1B	603	76%	890	#######
Chicken_VTR2A				
Chicken_VTR2B				
Chicken_VTR2C	57.2	66%	244	3,00E-10
Frog_OTR	158	75%	471	7,00E-40
Frog_VTR1A	161	65%	384	8,00E-42
Frog_VTR1B	125	64%	310	4,00E-31
Frog_VTR2A	75.2	70%	517	8,00E-15
Frog_VTR2B				
Frog_VTR2C	46.4	81%	256	6,00E-07
Spotted gar_OTR	100	76%	475	6,00E-23
Spotted gar_VTR1A	120	71%	378	3,00E-29
Spotted gar_VTR1B				
Spotted gar_VTR2A	114	75%	271	7,00E-28
Spotted gar_VTR2B	71.6	66%	317	5,00E-15
Spotted gar_VTR2C				
Coelacanth_OTR	59	69%	215	2,00E-10
Coelacanth_VTR1A	69.8	72%	210	2,00E-14
Coelacanth_VTR1B	167	64%	386	2,00E-43
Coelacanth_VTR2A	123	74%	602	3,00E-29
Coelacanth_VTR2B	77	74%	1,00E-16	262
Coelacanth_VTR2C				
Elephant shark_OTR	127	72%	357	3,00E-31
Elephant shark_VTR1A	187	72%	636	3,00E-49
Elephant shark_VTR1B	383	70%	820	#######
Elephant shark_VTR2A				
Elephant shark_VTR2B	84.2	68%	353	8,00E-19
Elephant shark_VTR2C	59	65%	249	2,00E-11
Sea lamprey_OTR	179	67%	409	5,00E-47
Sea lamprey_VTR1A	205	68%	551	3,00E-54
Sea lamprey_VTR1B				
Sea lamprey_VTR2A				
Sea lamprey_VTR2B	82.4	78%	430	8,00E-18
Sea lamprey_VTR2C	152	74%	333	1,00E-38

Human_VTR2A	Max Score	Identity	Total score	E-value
Chicken_OTR	91.5	65%	377	4,00E-21
Chicken_VTR1A	120	73%	890	7,00E-30
Chicken_VTR1B	123	76%	686	5,00E-31
Chicken_VTR2A				
Chicken_VTR2B				
Chicken_VTR2C	51.8	78%	133	8,00E-09
Frog_OTR	73.4	69%	657	1,00E-14
Frog_VTR1A	96.9	71%	307	2,00E-22
Frog_VTR1B	48.2	83%	430	6,00E-08
Frog_VTR2A	95.1	70%	433	5,00E-21
Frog_VTR2B				
Frog_VTR2C	48.2	79%	239	1,00E-07
Spotted gar_OTR	55.4	87%	338	1,00E-09
Spotted gar_VTR1A	87.8	69%	295	1,00E-19
Spotted gar_VTR1B				
Spotted gar_VTR2A	147	73%	596	7,00E-38
Spotted gar_VTR2B	69.8	76%	344	1,00E-14
Spotted gar_VTR2C	48.2	73%	303	1,00E-07
Coelacanth_OTR	44.6	82%	246	2,00E-06
Coelacanth_VTR1A	59	71%	260	3,00E-11
Coelacanth_VTR1B	46.4	78%	162	3,00E-07
Coelacanth_VTR2A	158	68%	556	9,00E-40
Coelacanth_VTR2B	48.2	70%	354	4,00E-08
Coelacanth_VTR2C				
Elephant shark_OTR	82.4	72%	525	6,00E-18
Elephant shark_VTR1	80.6	69%	396	2,00E-17
Elephant shark_VTR1	105	75%	689	2,00E-25
Elephant shark_VTR2				
Elephant shark_VTR2	93.3	67%	418	1,00E-21
Elephant shark_VTR2	53.6	75%	350	6,00E-10
Sea lamprey_OTR	111	72%	389	1,00E-26
Sea lamprey_VTR1A	140	76%	449	6,00E-35
Sea lamprey_VTR1B				
Sea lamprey_VTR2A				
Sea lamprey_VTR2B	78.8	69%	591	6,00E-17
Sea lamprey_VTR2C	161	71%	754	1,00E-41

Chicken_OTR	Max Score	Identity	Total score	E-value
Human_OTR	682	78%	1280	0.0
Human_VTR1A	179	65%	301	2,00E-47
Human_VTR1B	250	67%	653	1,00E-68
Human_VTR2A	91.5	65%	377	4,00E-21
Human_VTR2B				
Human_VTR2C				
Frog_OTR	646	77%	991	0.0
Frog_VTR1A	123	64%	265	9,00E-31
Frog_VTR1B	73.4	72%	232	1,00E-15
Frog_VTR2A	51.8	71%	193	4,00E-08
Frog_VTR2B				
Frog_VTR2C				
Spotted gar_OTR	419	71%	669	#######
Spotted gar_VTR1A	215	68%	500	2,00E-58
Spotted gar_VTR1B				
Spotted gar_VTR2A	91.5	72%	520	3,00E-21
Spotted gar_VTR2B	84.2	71%	195	4,00E-19
Spotted gar_VTR2C				
Coelacanth_OTR	343		539	1,00E-96
Coelacanth_VTR1A	82.4	69%	231	2,00E-18
Coelacanth_VTR1B	73.4	81%	181	2,00E-15
Coelacanth_VTR2A	141	79%	644	4,00E-35
Coelacanth_VTR2B	75.2	70%	117	2,00E-16
Coelacanth_VTR2C				
Elephant shark_OTR	497	73%	781	#######
Elephant shark_VTR1A	242	67%	328	2,00E-66
Elephant shark_VTR1B	212	68%	421	1,00E-57
Elephant shark_VTR2A				
Elephant shark_VTR2B	100	72%	211	5,00E-24
Elephant shark_VTR2C	96.9	72%	181	4,00E-23
Sea lamprey_OTR	381	70%	520	#######
Sea lamprey_VTR1A	233	71%	446	2,00E-63
Sea lamprey_VTR2B	73.4	80%	368	2,00E-15
Sea lamprey_VTR2C	159	77%	234	3,00E-41

Chicken_VTR1A	Max Score	Identity	Total score	E-value
Human_OTR	329	71%	781	4,00E-92
Human_VTR1A	738	80%	1006	0.0
Human_VTR1B	279	72%	1211	2,00E-77
Human_VTR2A	120	73%	890	7,00E-30
Human_VTR2B				
Human_VTR2C				
Frog_OTR	120	75%	359	6,00E-29
Frog_VTR1A	360	70%	796	#######
Frog_VTR1B	77	71%	449	7,00E-17
Frog_VTR2A	77	69%	452	8,00E-16
Frog_VTR2B				
Frog_VTR2C	53.6	84%	564	1,00E-09
Spotted gar_OTR	95.1	72%	305	9,00E-22
Spotted gar_VTR1A	403	71%	917	#######
Spotted gar_VTR1B				
Spotted gar_VTR2A	96.9	73%	672	6,00E-23
Spotted gar_VTR2B	84.2	69%	343	3,00E-19
Spotted gar_VTR2C	68	68%	255	8,00E-14
Coelacanth_OTR	82.4	80%	171	5,00E-18
Coelacanth_VTR1A	165	66%	456	1,00E-43
Coelacanth_VTR1B	80.6	89%	477	9,00E-18
Coelacanth_VTR2A	107	73%	816	8,00E-25
Coelacanth_VTR2B	62.6	82%	426	9,00E-13
Coelacanth_VTR2C				
Elephant shark_OTR	181	66%	470	5,00E-48
Elephant shark_VTR1A	347	70%	960	5,00E-98
Elephant shark_VTR1B	329	72%	716	4,00E-93
Elephant shark_VTR2A				
Elephant shark_VTR2B	111	71%	444	2,00E-27
Elephant shark_VTR2C	51.8	68%	417	1,00E-09
Sea lamprey_OTR	338	70%	625	3,00E-95
Sea lamprey_VTR1A	358	74%	958	#######
Sea lamprey_VTR2B	127	73%	656	7,00E-32
Sea lamprey_VTR2C	150	76%	465	1,00E-38

Chicken_VTR1B	Max Score	Identity	Total score	E-value
Human_OTR	210	67%	694	2,00E-56
Human_VTR1A	282	69%	459	2,00E-78
Human_VTR1B	603	76%	890	3,00E-175
Human_VTR2A	123	76%	686	5,00E-31
Human_VTR2B				
Human_VTR2C				
Frog_OTR	84.2	84%	597	3,00E-18
Frog_VTR1A	136	67%	325	9,00E-35
Frog_VTR1B	138	66%	329	2,00E-35
Frog_VTR2A				
Frog_VTR2B				
Frog_VTR2C	53.6	75%	74.7	1,00E-09
Spotted gar_OTR	100	78%	337	2,00E-23
Spotted gar_VTR1A	120	65%	359	8,00E-30
Spotted gar_VTR1B				
Spotted gar_VTR2A	78.8	89%	480	1,00E-17
Spotted gar_VTR2B	57.2	73%	57.2	3,00E-11
Spotted gar_VTR2C				
Coelacanth_OTR	68	87%	251	8,00E-14
Coelacanth_VTR1A	53.6	76%	157	5,00E-10
Coelacanth_VTR1B	152	66%	508	1,00E-39
Coelacanth_VTR2A	93.3	72%	249	1,00E-20
Coelacanth_VTR2B	66.2	80%	125	6,00E-14
Coelacanth_VTR2C				
Elephant shark_OTR	111	65%	346	5,00E-27
Elephant shark_VTR1A	145	67%	429	3,00E-37
Elephant shark_VTR1B	306	71%	600	4,00E-86
Elephant shark_VTR2A				
Elephant shark_VTR2B	73.4	67%	228	4,00E-16
Elephant shark_VTR2C				
Sea lamprey_OTR	152	67%	429	2,00E-39
Sea lamprey_VTR1A	280	71%	623	1,00E-77
Sea lamprey_VTR2B	80.6	70%	517	7,00E-18
Sea lamprey_VTR2C	80.6	67%	316	2,00E-17

Chicken_VTR2C	Max Score	Identity	Total score	E-value
Human_OTR				
Human_VTR1A	60.8	67%	602	3,00E-11
Human_VTR1B	57.2	66%	244	3,00E-10
Human_VTR2A	51.8	78%	133	8,00E-09
Human_VTR2B				
Human_VTR2C				
Frog_OTR	98.7	72%	591	5,00E-22
Frog_VTR1A	73.4	73%	334	3,00E-15
Frog_VTR1B	75.2	64%	403	6,00E-16
Frog_VTR2A				
Frog_VTR2B				
Frog_VTR2C	475	74%	782	#######
Spotted gar_OTR	98.7	67%	222	2,00E-22
Spotted gar_VTR1A	100	71%	429	3,00E-23
Spotted gar_VTR1B				
Spotted gar_VTR2A	42.8	75%	95.8	3,00E-06
Spotted gar_VTR2B	185	66%	444	3,00E-49
Spotted gar_VTR2C	385	71%	1001	#######
Coelacanth_OTR	69.8	70%	278	9,00E-14
Coelacanth_VTR1A	55.4	65%	179	5,00E-10
Coelacanth_VTR1B				
Coelacanth_VTR2A	107	68%	371	2,00E-24
Coelacanth_VTR2B	224	67%	336	3,00E-61
Coelacanth_VTR2C				
Elephant shark_OTR	102	74%	372	1,00E-23
Elephant shark_VTR1A	28.3	90%	270	0.18
Elephant shark_VTR1B				
Elephant shark_VTR2A				
Elephant shark_VTR2B	237	68%	482	6,00E-65
Elephant shark_VTR2C	167	68%	341	6,00E-44
Sea lamprey_OTR	46.4	76%	515	7,00E-07
Sea lamprey_VTR1A	42.8	90%	349	2,00E-05
Sea lamprey_VTR2B	48.2	79%	202	2,00E-07
Sea lamprey_VTR2C	138	73%	327	2,00E-34

Sea lamprey_OTR					Sea lamprey_VTR1A					Sea lamprey_VTR2B					Sea lamprey_VTR2C				
	Max Score	Identity	Total score	E-value		Max Score	Identity	Total score	E-value		Max Score	Identity	Total score	E-value		Max Score	Identity	Total score	E-value
Frog_OTR	210	68%	293	2,00E-55	Frog_OTR	129	73%	1033	1,00E-30	Frog_OTR	78.8	77%	408	5,00E-16	Frog_OTR	127	75%	347	2,00E-30
Frog_VTR1A	187	66%	322	3,00E-49	Frog_VTR1A	122	72%	563	2,00E-29	Frog_VTR1A					Frog_VTR1A	51.8	69%	104	2,00E-08
Frog_VTR1B	91.5	65%	236	1,00E-20	Frog_VTR1B	48.2	78%	407	3,00E-07	Frog_VTR1B	44.6	77%	173	1,00E-06	Frog_VTR1B	41	69%	205	3,00E-05
Frog_VTR2A	59	82%	195	7,00E-10	Frog_VTR2A	66.2	88%	412	1,00E-11	Frog_VTR2A	48.2	94%	241	1,00E-06	Frog_VTR2A	91.5	66%	288	2,00E-19
Frog_VTR2B					Frog_VTR2B					Frog_VTR2B					Frog_VTR2B				
Frog_VTR2C					Frog_VTR2C	55.4	79%	518	3,00E-09	Frog_VTR2C	86	78%	371	7,00E-19	Frog_VTR2C	98.7	66%	307	2,00E-22
Spotted gar_OTR	127	65%	305	5,00E-31	Spotted gar_OTR	107	73%	689	1,00E-24	Spotted gar_OTR	55.4	77%	232	2,00E-09	Spotted gar_OTR	77	70%	514	1,00E-15
Spotted gar_VTR1A	297	68%	536	2,00E-82	Spotted gar_VTR1A	181	75%	865	3,00E-47	Spotted gar_VTR1A	44.6	66%	233	2,00E-06	Spotted gar_VTR1A	86	70%	110	1,00E-18
Spotted gar_VTR1B					Spotted gar_VTR1B					Spotted gar_VTR1B					Spotted gar_VTR1B				
Spotted gar_VTR2A	77	83%	443	2,00E-16	Spotted gar_VTR2A	118	71%	898	2,00E-28	Spotted gar_VTR2A	82.4	74%	442	4,00E-18	Spotted gar_VTR2A	116	76%	322	4,00E-28
Spotted gar_VTR2B	73.4	68%	172	2,00E-15	Spotted gar_VTR2B	86	74%	728	7,00E-19	Spotted gar_VTR2B	138	69%	369	4,00E-35	Spotted gar_VTR2B	118	71%	416	8,00E-29
Spotted gar_VTR2C	53.6	75%	148	6,00E-09	Spotted gar_VTR2C	53.6	66%	1178	1,00E-08	Spotted gar_VTR2C	78.8	74%	421	1,00E-16	Spotted gar_VTR2C				
Coelacanth_OTR	93.3	66%	256	9,00E-21	Coelacanth_OTR	60.8	66%	1071	1,00E-10	Coelacanth_OTR					Coelacanth_OTR	41	70%	302	9,00E-05
Coelacanth_VTR1A	98.7	71%	192	6,00E-23	Coelacanth_VTR1A	57.2	73%	822	4,00E-10	Coelacanth_VTR1A					Coelacanth_VTR1A	69.8	77%	359	5,00E-14
Coelacanth_VTR1B	125	64%	311	8,00E-31	Coelacanth_VTR1B	84.2	77%	498	6,00E-18	Coelacanth_VTR1B					Coelacanth_VTR1B				
Coelacanth_VTR2A	111	69%	345	2,00E-25	Coelacanth_VTR2A	89.7	69%	907	2,00E-18	Coelacanth_VTR2A	75.2	72%	416	1,00E-14	Coelacanth_VTR2A	141	70%	402	2,00E-34
Coelacanth_VTR2B	68	79%	219	7,00E-14	Coelacanth_VTR2B	57.2	77%	573	3,00E-10	Coelacanth_VTR2B	69.8	72%	216	2,00E-14	Coelacanth_VTR2B	114	71%	297	1,00E-27
Coelacanth_VTR2C					Coelacanth_VTR2C					Coelacanth_VTR2C					Coelacanth_VTR2C				
Elephant shark_OTR	262	68%	643	6,00E-72	Elephant shark_OTR	158	75%	1063	4,00E-40	Elephant shark_OTR	68	76%	405	2,00E-13	Elephant shark_OTR	102	73%	153	2,00E-23
Elephant shark_VTR1A	329	74%	604	5,00E-92	Elephant shark_VTR1A	170	77%	1001	7,00E-44	Elephant shark_VTR1A	57.2	67%	410	4,00E-10	Elephant shark_VTR1A	113	72%	258	1,00E-26
Elephant shark_VTR1B	206	67%	280	1,00E-55	Elephant shark_VTR1B	165	69%	701	9,00E-43	Elephant shark_VTR1B	55.4	73%	257	4,00E-10	Elephant shark_VTR1B	73.4	65%	238	3,00E-15
Elephant shark_VTR2A					Elephant shark_VTR2A					Elephant shark_VTR2A					Elephant shark_VTR2A				
Elephant shark_VTR2B	82.4	68%	199	4,00E-18	Elephant shark_VTR2B	98.7	70%	791	1,00E-22	Elephant shark_VTR2B	127	68%	409	8,00E-32	Elephant shark_VTR2B	163	68%	322	2,00E-42
Elephant shark_VTR2C					Elephant shark_VTR2C	75.2	71%	340	9,00E-16	Elephant shark_VTR2C	87.8	79%	171	4,00E-20	Elephant shark_VTR2C	122	69%	339	5,00E-30

Table S6: BLASTn comparisons between human, chicken and sea lamprey OTR-VTR against all OTR-VTR present in human, chicken, frog, spotted gar, coelacanth, elephant shark and lamprey. Red highlighting denotes comparisons that yielded results below threshold (max. score<40, E-value > 10^{-4}). Black highlight indicates absence of this receptor in this species.

Table S7: Intraspecies (human) synteny analysis in the territory of OTR-VTROTR-VTR. Genes shaded with colors: genes found in a 10 Mb window surrounding the OTR-VTR of interest. Genes with an asterisk: Genes from gene families that have been independently found to reside in syntenic blocks (http://ohnologs.curie.fr/cgi-bin/BrowsePage.cgi?org=human). Genes in red: genes from larger superfamilies that were not included in the Figures. For gene family-symbols, we followed the nomenclature proposed by the HGNC (HUGO Gene Nomenclature Committee).

Family	"VTR2B" (Chr. 3)		"VTR2Ca" (Chr. 7)		"VTR2Cb" (Chr. 12)		VTR2A (Chr. X)	
ABCD *					ABCD2 *	39,5	ABCD1*	153,7
AMIGO *	AMIGO3 *	49,6			AMIGO2 *	47,0		
ARF GTPase	ARF4 *	57,5	ARF5 *	127,6	ARF3	48,9		
	ARL8B	5,1						
ASB	ASB14	57,2	ASB4	95,4	ASB8	48,1	ASB9	15,2
			ASB15	123,5				
CACNA1	CACNA1D	53,5			CACNA1C	1,9	CACNA1F	49,2
CACNA2D	CACNA2D2 *	50,4	CACNA2D1 *	81,6	CACNA2D4	1,9		
	CACNA2D3 *	54,2						
CCDC	CCDC66	56,5	CCDC51L	106,6	CCDC184	48,1		
	CCDC31	48,4			CCDC65	48,9		
	CCDC71	49,1						
	CCDC36	49,1						
CDHR *	CDHR4 *	49,7	CDHR3 *	105,8				
CNTN *	CNTN6 *	1,1	CNTNAP2	146,1	CNTN1 *	40,6		
	CNTN4 *	2,1						
	CNTN3 *	74,3						
COL *	COL7A1 *	48,5			COL2A1 *	47,9		
DENND *	DENND6A	57,6	DENDD2A *	140,5	DENND5B	31,3		
DNASE	DNASE1L3	58,1					DNASE1L1	154,4
DNAJ			DNAJC2	103,3	DNAJC22	49,3		
			DNAJB9	108,5				
DOCK *	DOCK3 *	50,6	DOCK4 *	111,7			DOCK11	118,4
DUSP * (MAP	DUSP7 *	52,0	DUSP24 (STY	75,9	DUSP6 *	89,3	DUSP9 *	153,6
FLN *	FLNB *	58,0	FLNC *	128,5			FLNA *	153,6
GNAI *	GNAI2 *	50,3	GNAI1 *	79,8				
GNAT *	GNAT1 *	50,2	GNAT3 *	80,1				
GPR *	GPR62 *	51,9	GPR22	107,4			GPR50	151,1
			GPR85	112,9				
GRM *	GRM2 *	51,7	GRM3 *	86,6				
	GRM7 **	6,8	GRM8 **	126,4				
GXYLT *	GXYLT2 *	72,9			GXYLT1 *	42,5		
IFRD	IFRD2 *	50,2	IFRD1 *	112,4				
ILR	IL5RA	3,1					IL9R	155,9
	IL17RB	53,8						
	IL17RD	57,0						
IRAK *					IRAK4 *	43,7	IRAK1 *	154,0
KMT			KMT2E	105	KMT2D	49		
L1CAM *	CHL1 *	0,2	NRCAM *	107,8			L1CAM *	153,1
LAMB *	LAMB2 *	49,1	LAMB1 *	107,9				107,9
			LAMB4 *	108,0				
LRRN *	LRRN1 *	3,8	LRRN3 *	110,7				
LSMEM	LSMEM2	50,2	LSMEM1	112,4				
MAGI *	MAGI1 *	65,3	MAGI2 *	77,6			MAGIX *	49,1
MTND pseudo					MTND1P24 g	41,6		
					MTND2P17	41,6		
OPN *	RHO (OPN2)	129,2	OPN1SW *	128,4			OPN1LW *	153,4
							OPN1MW *	153,4
							OPN1MW2 *	153,5
OT/VP-R	"VTR2B"-deleted		"V2TRCa"-deleted		"VTR2Cb"-deleted		VTR2A	152,8
PLXNA *	PLXNA1 *	126,7	PLXNA4 *	131,8			PLXNA3 *	
PLXNB *	PLXNB1 *	48,4					PLXNB3 *	153,7
PPP1R			PPP1R3A	113,8	PFKM (PPP1	48,1		153,7
PPM1 *	PPM1M *	52,3			PPM1H *	63,0		
PRICKLE *	PRICKLE2 *	64,1			PRICKLE1 *	42,4	PRICKLE3 *	
PTPN *	PTPN23	47,3	PTPN12 *	77,2				49,0
PUF			PUF60	105,4	PUF60	43,7		
RBM *	RBM15B *	51,4					RBM10 **	
	RBM6 **	49,9						47,1
	RBM5 **	50,0						
RHO	RHOA	49,3			RND1	48,8		
RPL	RPL29	51,9			RPL30P13 ps	40	RPL10	
					RPL21P101 ps	43,5		154,3
SEMA3	SEMA3F	50,2	SEMA3C	80,4				
	SEMA3B	50,3	SEMA3A	83,6				
	SEMA3G	52,5	SEMA3E	83,0				
			SEMA3D	84,6				
SLC *					SLC2A13	39,7	SLC6A8	
	SLC25A20	48,8			SLC38A1	46,1		153,6
					SLC38A2	46,3	SLC10A3	
	SLC26A6 *	48,6	SLC26A4 *	107,6	SLC38A4	46,7		154,4
	SLC26A5 *	103,3	SLC26A3 *	107,7	SLC48A1	47,7		
SYP	SYNPR	63,2	SYPL1	105,7			SYP	
SRPK			SRPK2 *	105,0			SRPK3 *	49,0
TMEM	TMEM115	50,3	TMEM168	112,7	TMEM117	43,7	TMEM187	153,7
					TMEM106C	47,9	TMEM185A	153,9
TREX	TREX1	48,4					TREX2	
TVP *	TVP23 *	52,3			TVP1 *	43,7		153,4
UBA*	UBA7*	49,8					UBA1*	
	UBA3	69,0						47,1
USP*	USP19	49,1			USP5	6,8	USP26	
	USP4*	49,3					USP11*	133,0
WNT*	WNT5A*	55,4	WNT2*	117,2	WNT5B*	1,5		47,1
			WNT16	121,2	WNT10B	48,9		
					WNT1	48,9		

Table S8: Intraspecies (human) synteny analysis in the territory of VTR2A and the territory of the deleted in the human genome VTR2B and VTR2C. The surrounding territory of VTR2B in spotted gar and of VTR2C in chicken were used to identify the territory of the lost VTR2B (Chr. 3:0-7 Mb and Chr.3:48-58 Mb) and VTR2C (VTR2Ca: Chr. 7:103-113 Mb and VTR2Cb: Chr. 12:39,5-50 Mb) in human. Genes shaded with colors: genes found in a 10 Mb window surrounding the OTR-VTR of interest. Genes not shaded: Genes found outside of the

strict 10 Mb window, but on the same chromosome as the OTR-VTR of interest. *Genes from gene families that have been independently found to reside in syntenic blocks (http://ohnologs.curie.fr/cgi-bin/BrowsePage.cgi?org=human). Genes in red: genes from larger superfamilies that were not included in the Figures. For gene family-symbols, we followed the nomenclature proposed by the HGNC (HUGO Gene Nomenclature Committee).

Studies	VTR1A	VTR1B	OTR	VTR2A	VTR2B	VTR2C
Nakatani et al. 2007	D	D	D	?	D	D
Putnam et al. 2008	13	13	13	13	13	13
Smith et al. 2013	AncD	AncD	AncD	AncD	AncD	AncD
Smith et al. 2018	5	5	5	5	5	5

Table S9: Mapping of OTR-VTR regions to putative ancestral chromosomes suggested in different studies. In Nakatani et al. 2007 the region of VTR2A (and a ~2Mb region around it) was not included in the analysis due to ambiguous synteny. In Smith et al. 2018, the presumptive ancestral chromosomes were not enumerated, but shaded with different colors in Figure 4 of their study. Our regions of interest go back to the 5th shaded pink region in their Figure 4.

PMZ_0045207-RA	scaf_00049:3426470..3434237				
	Max score	Total score	Query cover	E value	Ident
OTR	179	903	11%	5,00E-47	77%
VTR2B	262	1063	10%	4,00E-72	87%
VTR1A	152	762	8%	2,00E-38	73%
VTR2C	57.2	364	2%	7,00E-10	84%
PMZ_0014716-RA	**scaf_00049:4609394..4620022**				
	Max score	Total score	Query cover	E value	Ident
OTR	607	3379	23%	4,00E-175	85%
VTR2B	114	1506	6%	2,00E-27	96%
VTR1A	170	1093	5%	6,00E-44	82%
VTR2C	136	1288	9%	7,00E-34	70%

Table S10: Blastn comparisons between the putative OTR-VTR in sea lamprey scaffold 49 and the rest of the sea lamprey OTR-VTR.

Supplementary Figures

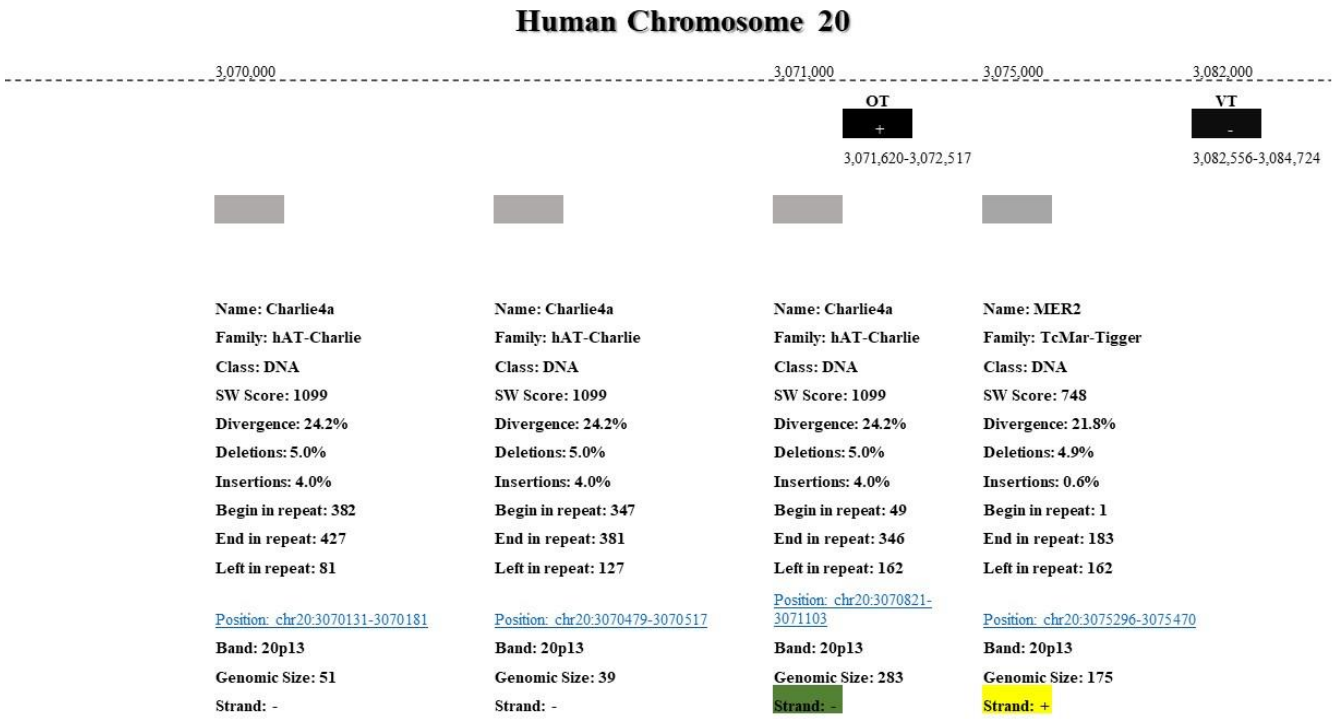


Figure S1: Illustration of OT and VT genes in the human genome, along with all the DNA transposable elements (TEs) flanking their territory. + and – denote orientation. The information on each DNA TE was retrieved from the UCSC Genome Browser database. According to Dfam, MER2 (MEdium Reiteration frequency interspersed) repeats are identified in the opposite orientation and are responsible for Target Site Duplications.

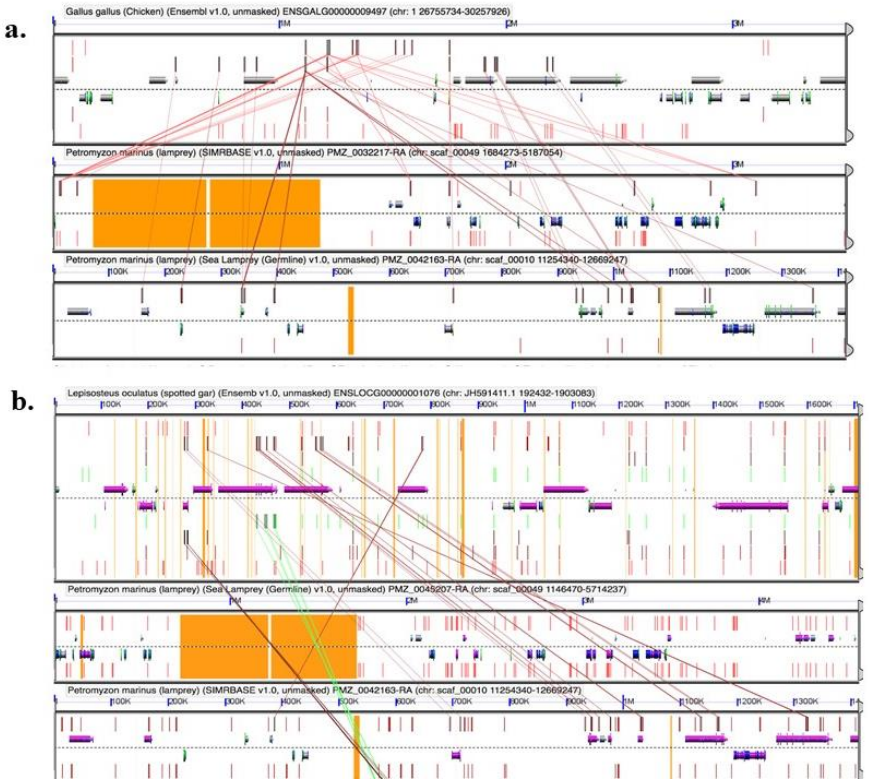
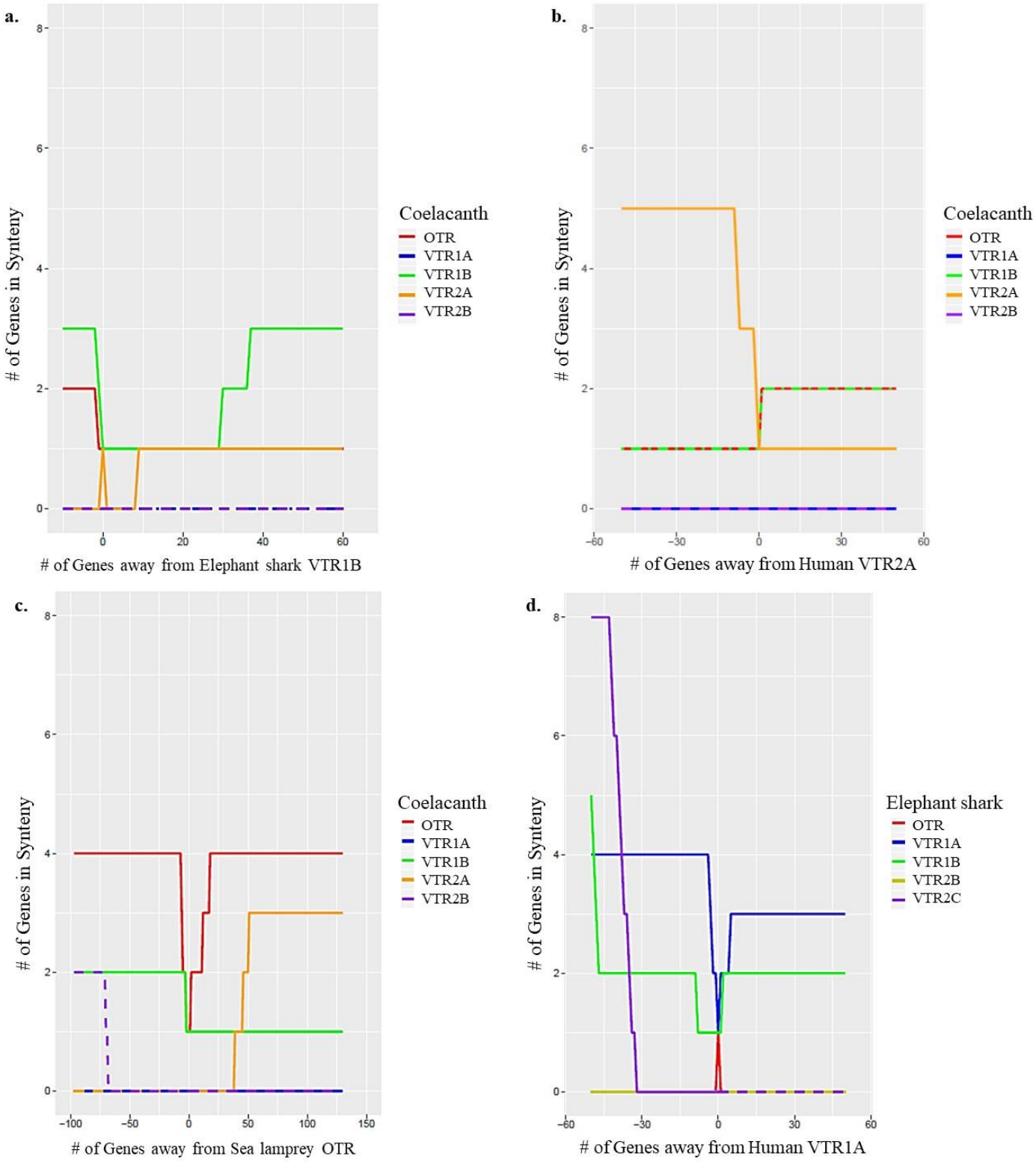
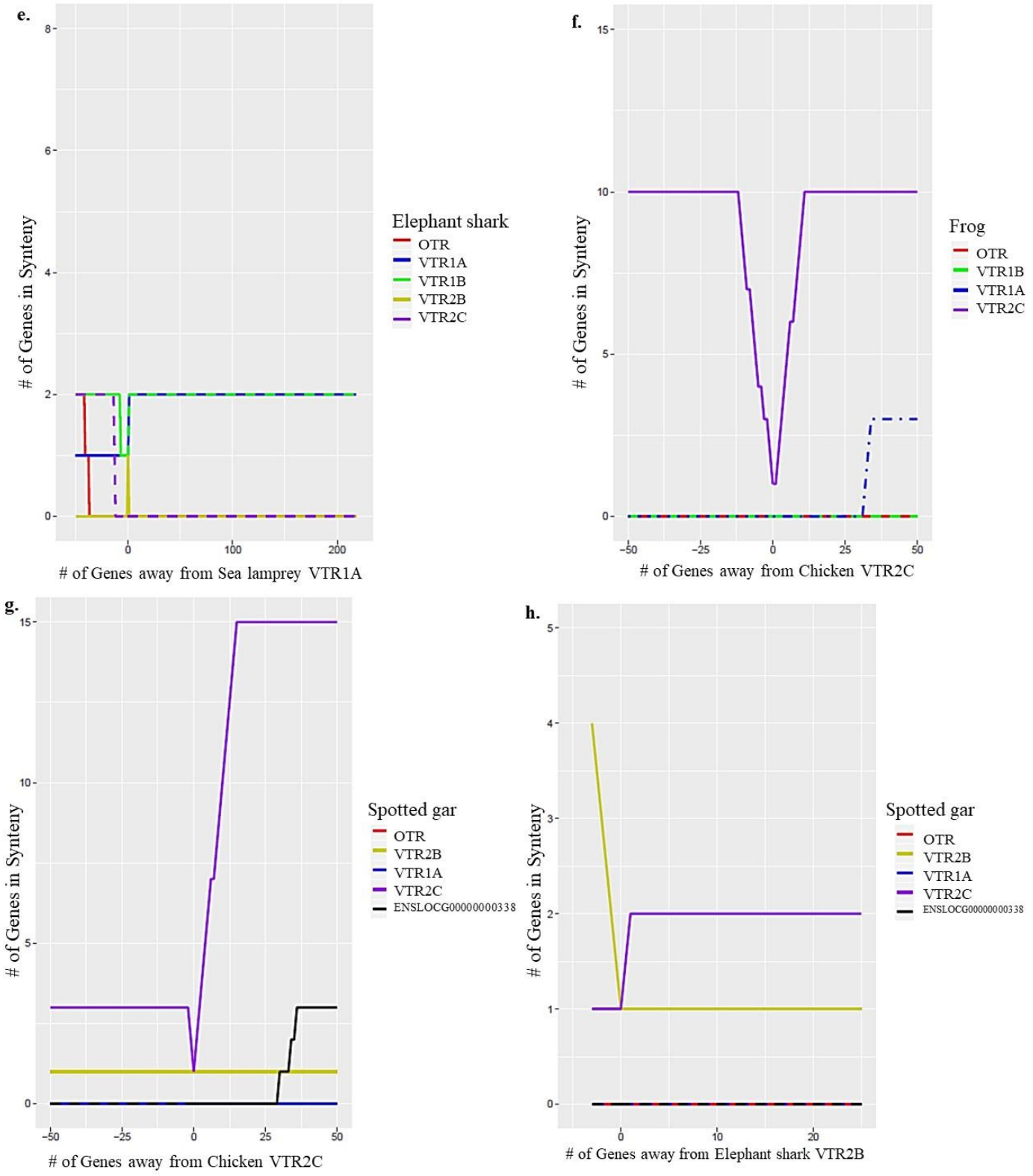


Figure S2: SynFind results illustrated by GeVo. a. Chicken VTR2C presents synteny (equal syntenic depth: https://genomevolution.org/r/10ukv) with both sea lamprey PMZ_0032217-RA (scaffold 49) and PMZ_0042163-RA (scaffold 10), which is the putative sea lamprey VTR2C-ortholog. b. Spotted gar VTR2C shows more synteny (syntenic depth: 7, https://genomevolution.org/r/10vin) with the PMZ_0045207-RA territory (scaffold 49), rather than with the PMZ_0042163-RA territory (scaffold 10) (syntenic depth: 3). Note: PMZ_0032217-RA and PMZ_0045207-RA are different transcripts of the same gene in sea lamprey.


a.
# of Genes in Synteny
# of Genes away from Elephant shark VTR1B
Coelacanth
OTR
VTR1A
VTR1B
VTR2A
VTR2B
b.
# of Genes in Synteny
# of Genes away from Human VTR2A
Coelacanth
OTR
VTR1A
VTR1B
VTR2A
VTR2B
c.
# of Genes in Synteny
# of Genes away from Sea lamprey OTR
Coelacanth
OTR
VTR1A
VTR1B
VTR2A
VTR2B
d.
# of Genes in Synteny
# of Genes away from Human VTR1A
Elephant shark
OTR
VTR1A
VTR1B
VTR2B
VTR2C


e.
# of Genes in Synteny
# of Genes away from Sea lamprey VTR1A
Elephant shark
OTR
VTR1A
VTR1B
VTR2B
VTR2C
f.
# of Genes in Synteny
# of Genes away from Chicken VTR2C
Frog
OTR
VTR1B
VTR1A
VTR2C
g.
# of Genes in Synteny
# of Genes away from Chicken VTR2C
Spotted gar
OTR
VTR2B
VTR1A
VTR2C
ENSLOCG00000000338
h.
# of Genes in Synteny
# of Genes away from Elephant shark VTR2B
Spotted gar
OTR
VTR2B
VTR1A
VTR2C
ENSLOCG00000000338

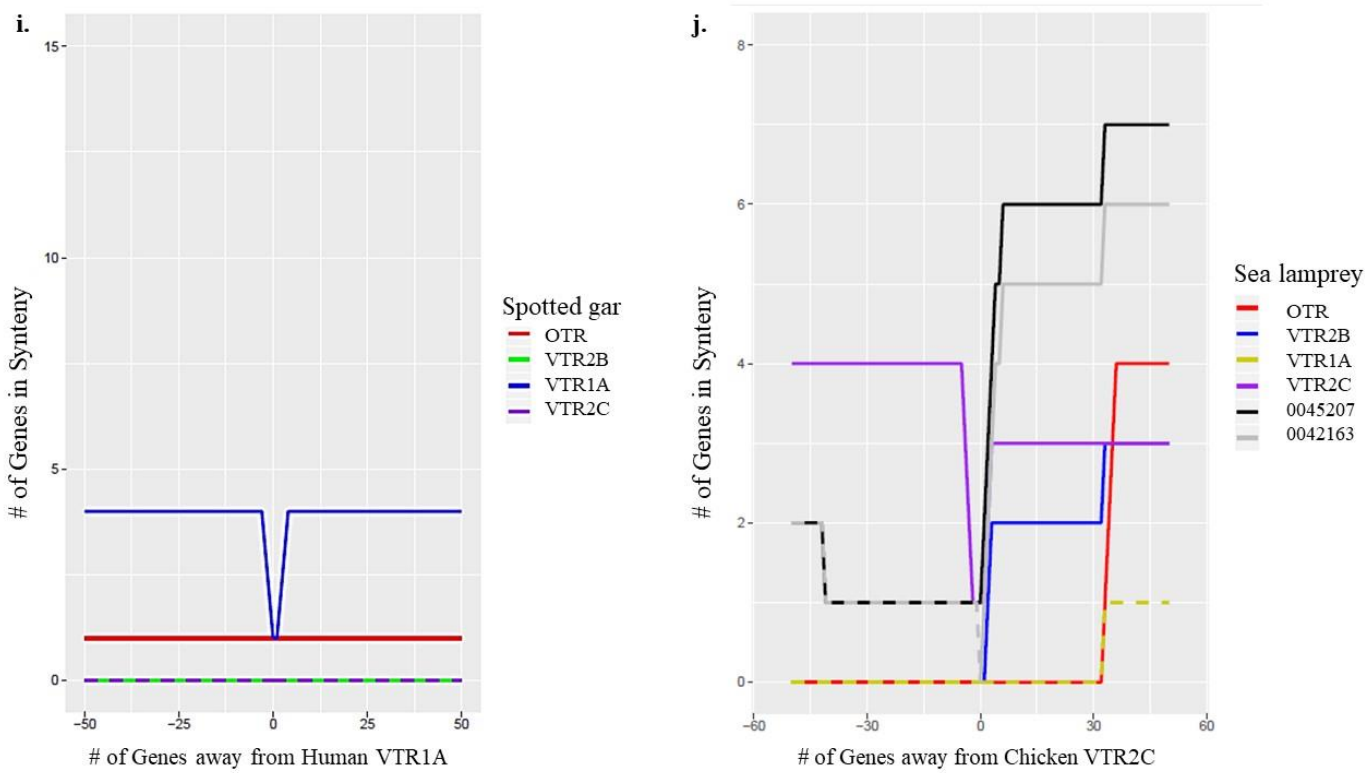


Figure S3 (a-j): Plots generated with SynFind: In the x axis 0 represents the query OTR-VTR in the query organism (e.g. VTR1B in elephant shark in a.) and the numbers represent the genes on the 5' (left) and on the 3' (right) of the query OTR-VTR in the genome (e.g. '30' genes on the left of '0' in b. represents the 30th gene on the left of VTR2A in the human genome). The y axis shows the number of the matched homologous genes in the reference genome for each reference receptor (e.g. in c. coelacanth OTR in red shows 4 syntenic gene matches with genes on the left of sea lamprey OTR, and 4 matches on the right of sea lamprey OTR). If the reference OTR-VTR does not show any match, then it stays in 0 in the y axis (e.g. in g. spotted gar VTR1A, in blue); if it matches only the query OTR-VTR it reaches 1 (e.g. in i. spotted gar OTR -in red- was only homologous to human VTR1A). If the reference OTR-VTR is not homologous to the query OTR-VTR but does show gene matches in the neighboring territory, then its line increases where the gene-match is located (e.g. in j. sea lamprey OTR -in red- hit 4 gene matches found approximately 30 genes on the right of chicken VTR2C, which is where its line increased and reached 4 in the y axis).

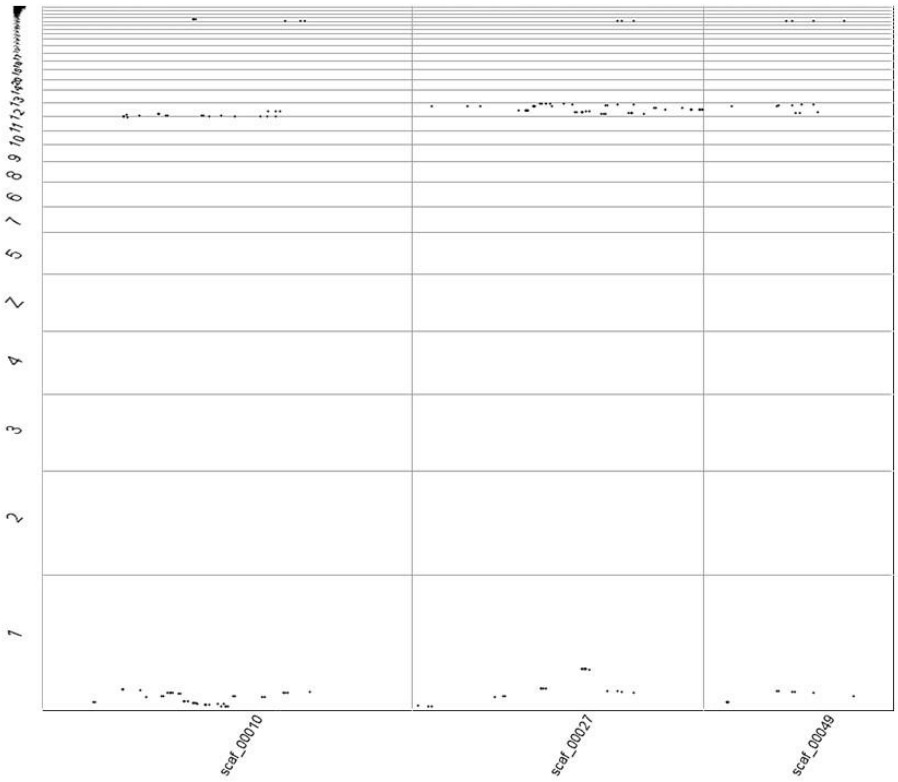


Figure S4: Example of a SynMap2-generated dotplot between sea lamprey scaffolds 10, 27 and 49 (x axis) and the chicken genome chromosomes (y axis). Each dot represents a syntenic homologous gene-hit between the two organisms (based on Last default parameters). Each gene-hit must be located at a maximum of a 20-gene distance from other two homologous gene-hits in each genome for it to be considered a syntenic gene-hit between these two species.

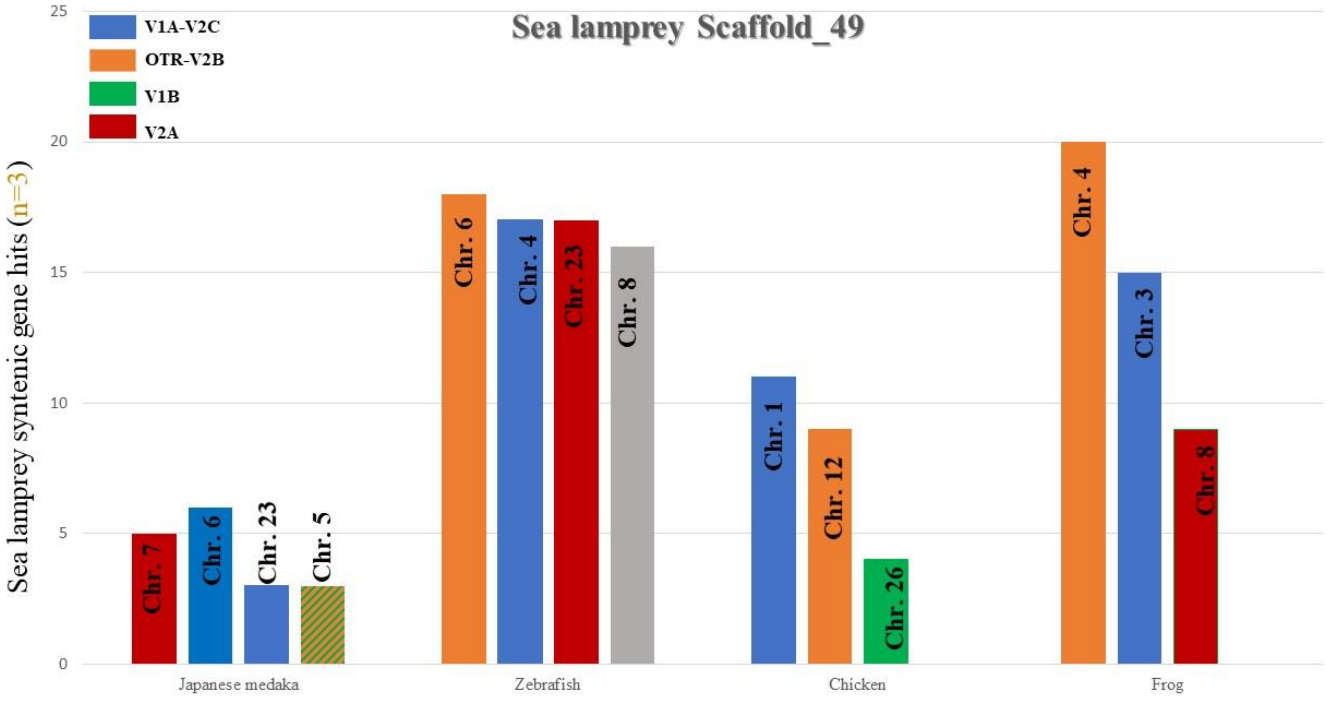


Figure S5: Number of syntenic gene hits between Japanese medaka, zebrafish, chicken and frog chromosomes against sea lamprey scaffold_49. * indicate that the difference in gene hits between the first two chromosomes

with the highest number of hits is statistically different ($p<0.05$; t test). n=3 means that the minimum number of aligned homologous gene pairs (at a maximum of a 20-gene distance from each other in each genome) for these genes to be considered syntenic was 3. Chromosomes are colored based on the OTR-VTR that resides in them.

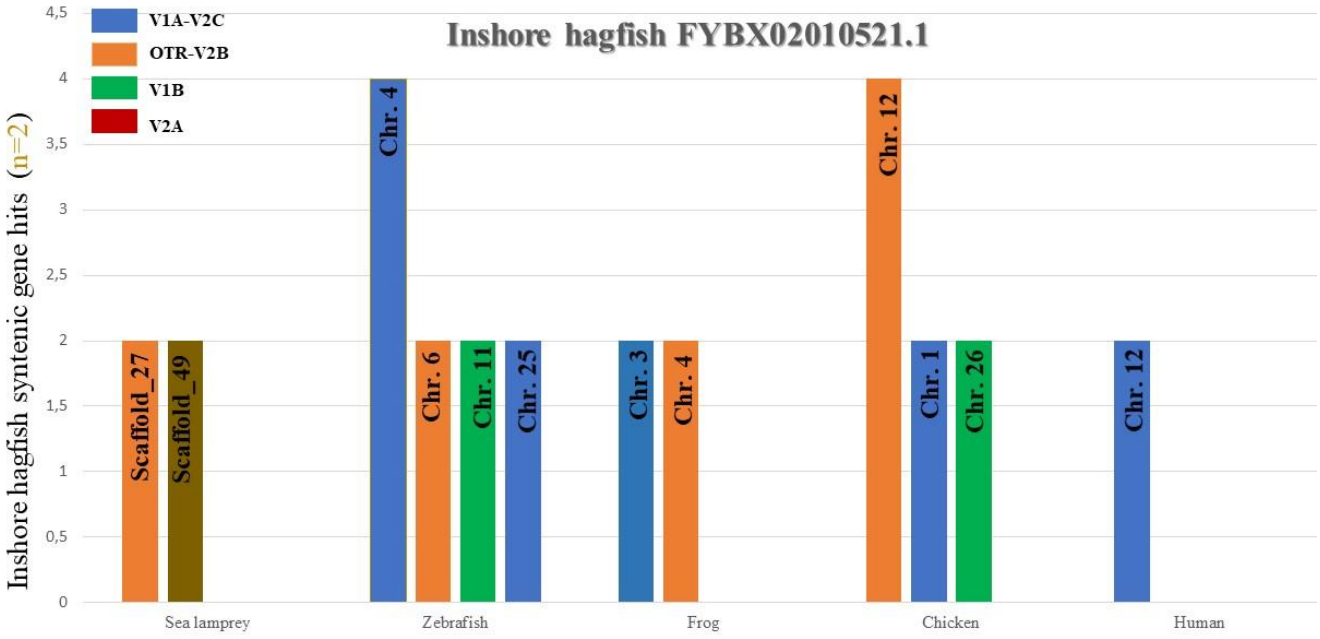


Figure S6: Number of syntenic gene hits between sea lamprey, zebrafish, frog, chicken and human chromosomes against inshore hagfish scaffold FYBX02010521.1. n=2 means that the minimum number of aligned homologous gene pairs (at a maximum of a 20-gene distance from each other in each genome) for these genes to be considered syntenic was 2. Chromosomes are colored based on the OTR-VTR that reside in them.

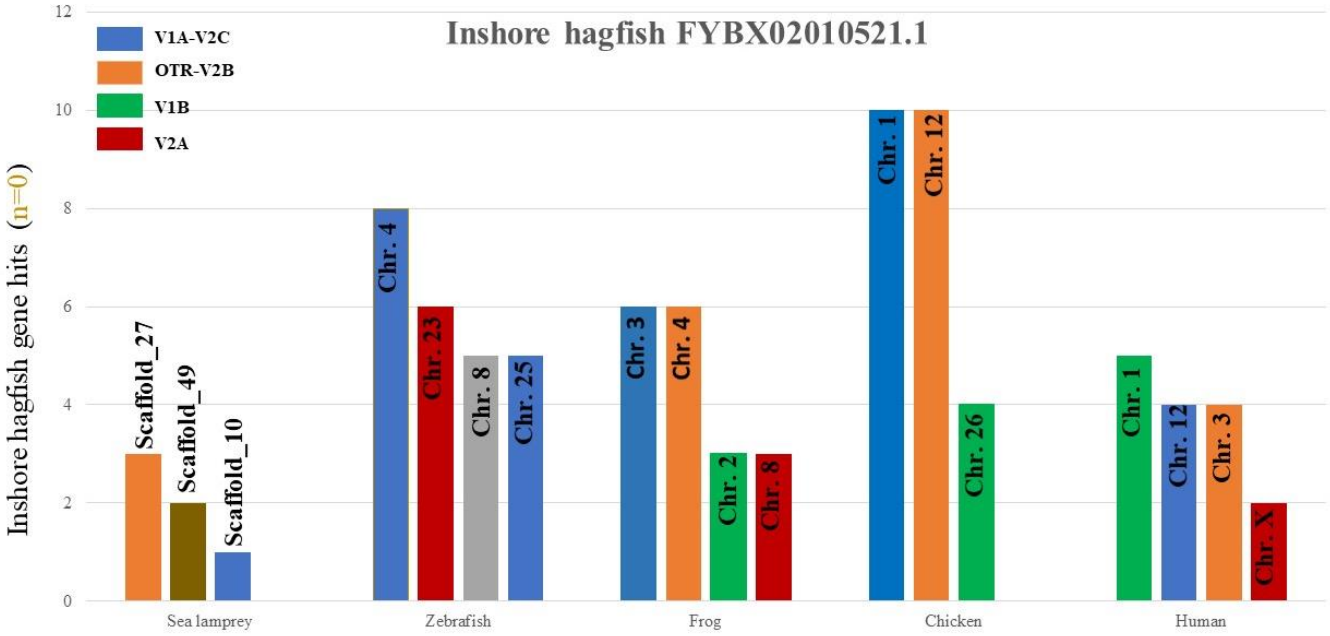


Figure S7: Number of homologous gene hits between sea lamprey, zebrafish, frog, chicken and human chromosomes against inshore hagfish scaffold FYBX02010521.1. n=0 means that in this analysis no synteny

was required for homologous gene pairs to appear in the results. All possible homologous gene pairs between chromosomes were shown in the analysis. Chromosomes are colored based on the OTR-VTR that reside in them.

Figure S8: Number of syntenic gene hits between sea lamprey, zebrafish, frog, chicken and human chromosomes against inshore hagfish scaffold FYBX02010841.1. n=2 means that the minimum number of aligned homologous gene pairs (at a maximum of a 20-gene distance from each other in each genome) for these genes to be considered syntenic was 3. Chromosomes are colored based on the OTR-VTR that reside in them.

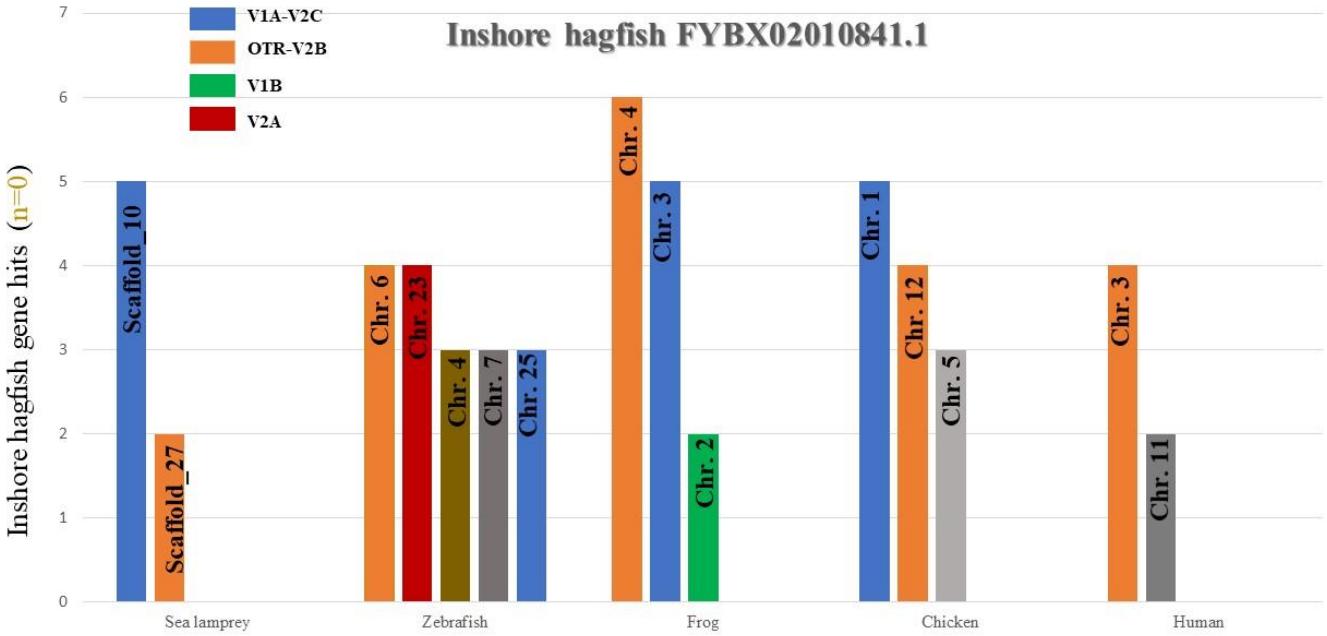


Figure S9: Number of homologous gene hits between sea lamprey, zebrafish, frog, chicken and human chromosomes against inshore hagfish scaffold FYBX02010841.1. n=0 means that in this analysis no synteny was required for homologous gene pairs to appear in the results. All possible homologous gene pairs between chromosomes were shown in the analysis. Chromosomes are colored based on the OTR-VTR that reside in them.

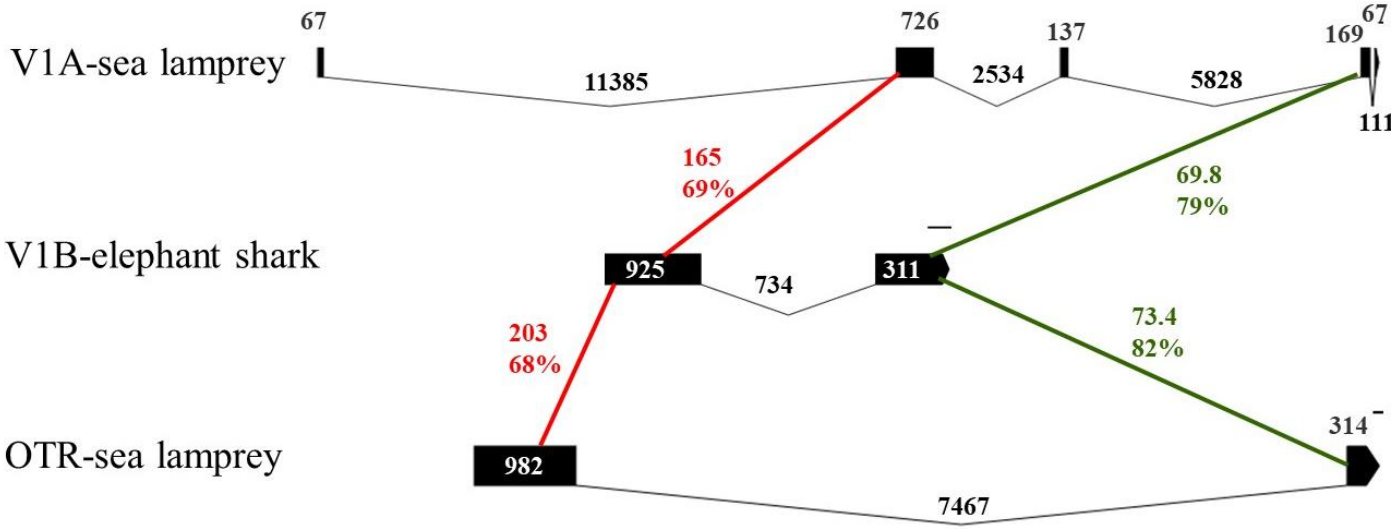


Figure S10: Comparison (via BLASTn) of the introns and exons of elephant shark VTR1B/V1B against introns and exons of sea lamprey VTR1A/V1A and sea lamprey OTR in all possible combinations. Max. scores and percent identities are shown for the alignments that yielded beyond threshold results (maximum score>40 and E-value< 10^{-4}). Sequence length is shown in bp.

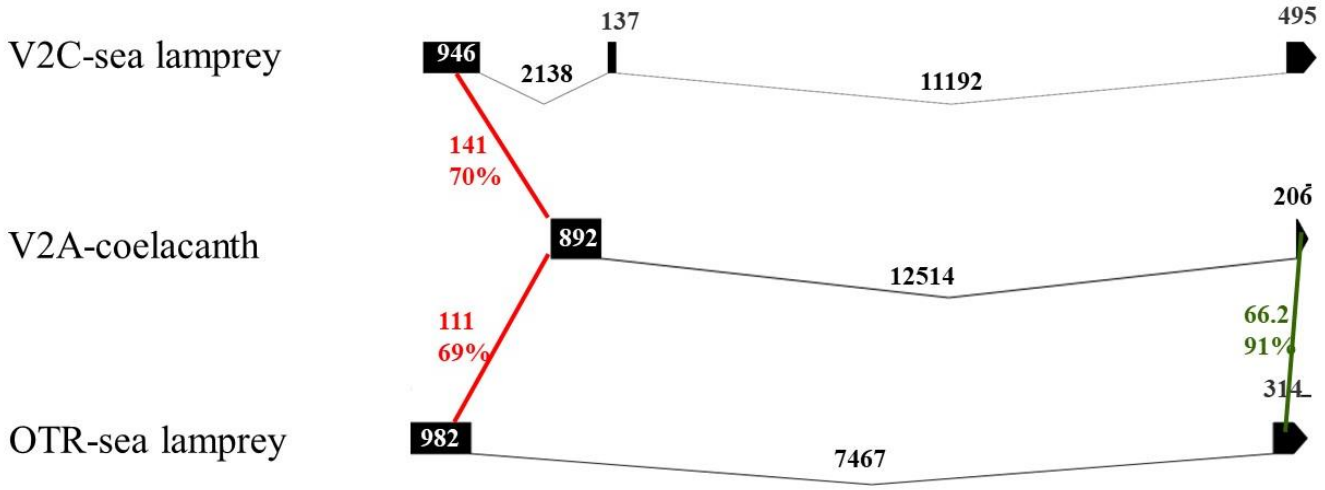


Figure S11: Comparison (via BLASTn) of the introns and exons of coelacanth VTR2A/V2A against introns and exons of sea lamprey VTR2C/V2C and sea lamprey OTR in all possible combinations. Max. scores and percent identities are shown for the alignments that yielded beyond threshold results (maximum score>40 and E-value< 10^{-4}). Sequence length is shown in bp.

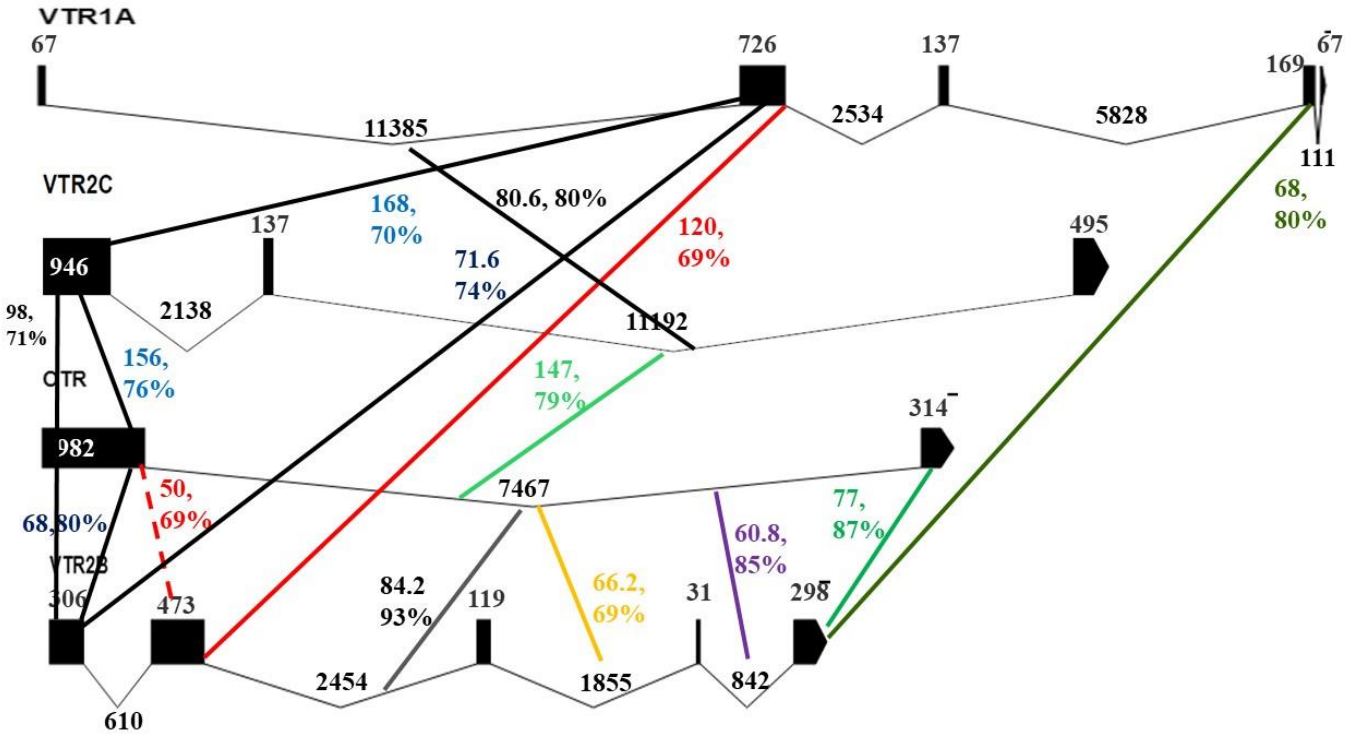


Figure S12: Comparisons in sea lamprey (via BLASTn) of the introns and exons of VTR1A against introns and exons of VTR2C and VTR2B; also comparisons of introns and exons of OTR against VTR2B in all possible combinations. Max. scores and percent identities are shown for the alignments that yielded beyond threshold results (maximum score>40 and E-value< 10^{-4}). Sequence length is shown in bp.

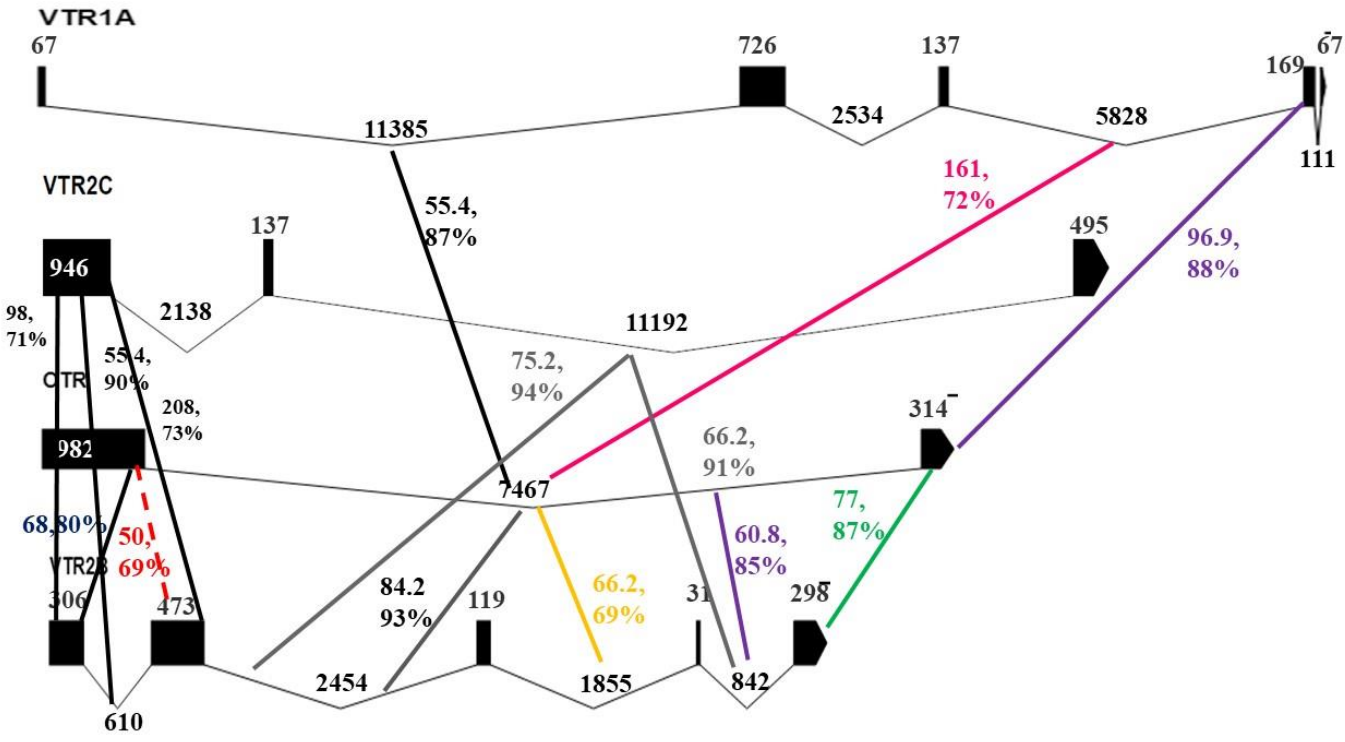


Figure S13: Comparisons in sea lamprey (via BLASTn) of the introns and exons of VTR2B against introns and exons of OTR and VTR2C; also comparisons of introns and exons of OTR against VTR2C in all possible combinations. Max. scores and percent identities are shown for the alignments that yielded beyond threshold results (maximum score>40 and E-value< 10^{-4}). Sequence length is shown in bp.

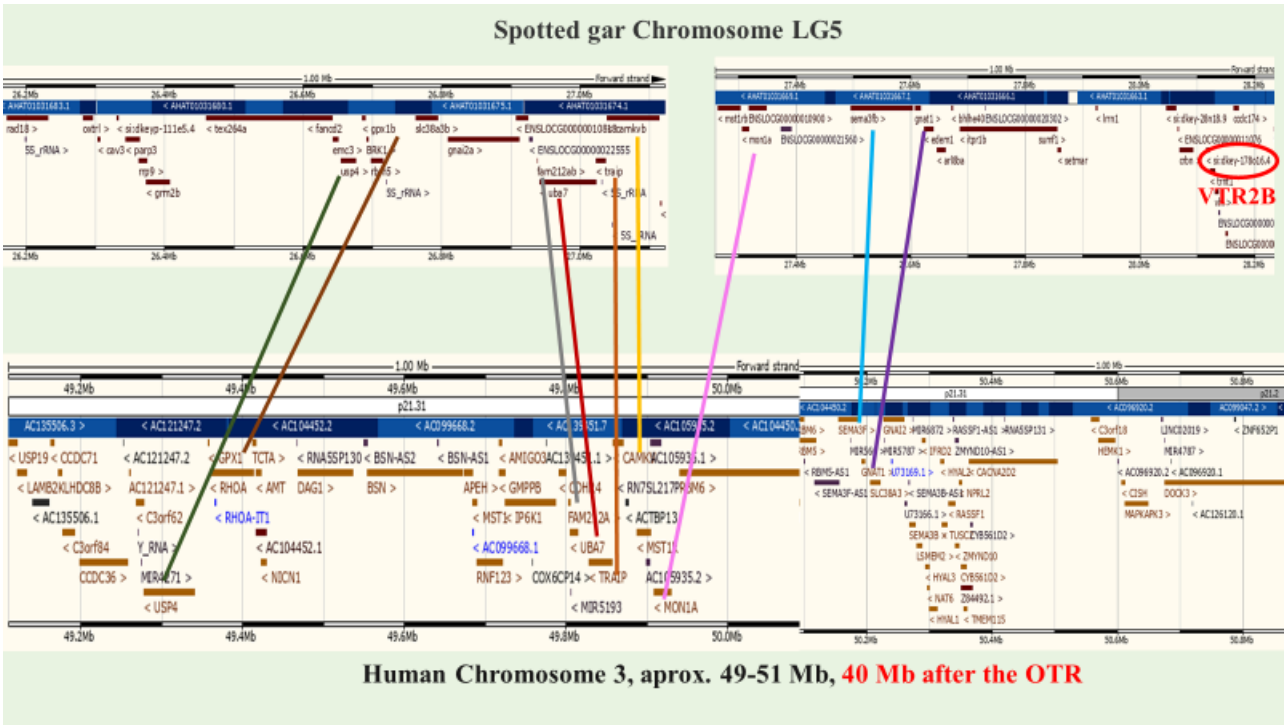


Figure S14: The genomic territory before the spotter gar VTR2B was found in human chromosome 3 (49-51 Mb), 40 Mb after the location of human OTR.

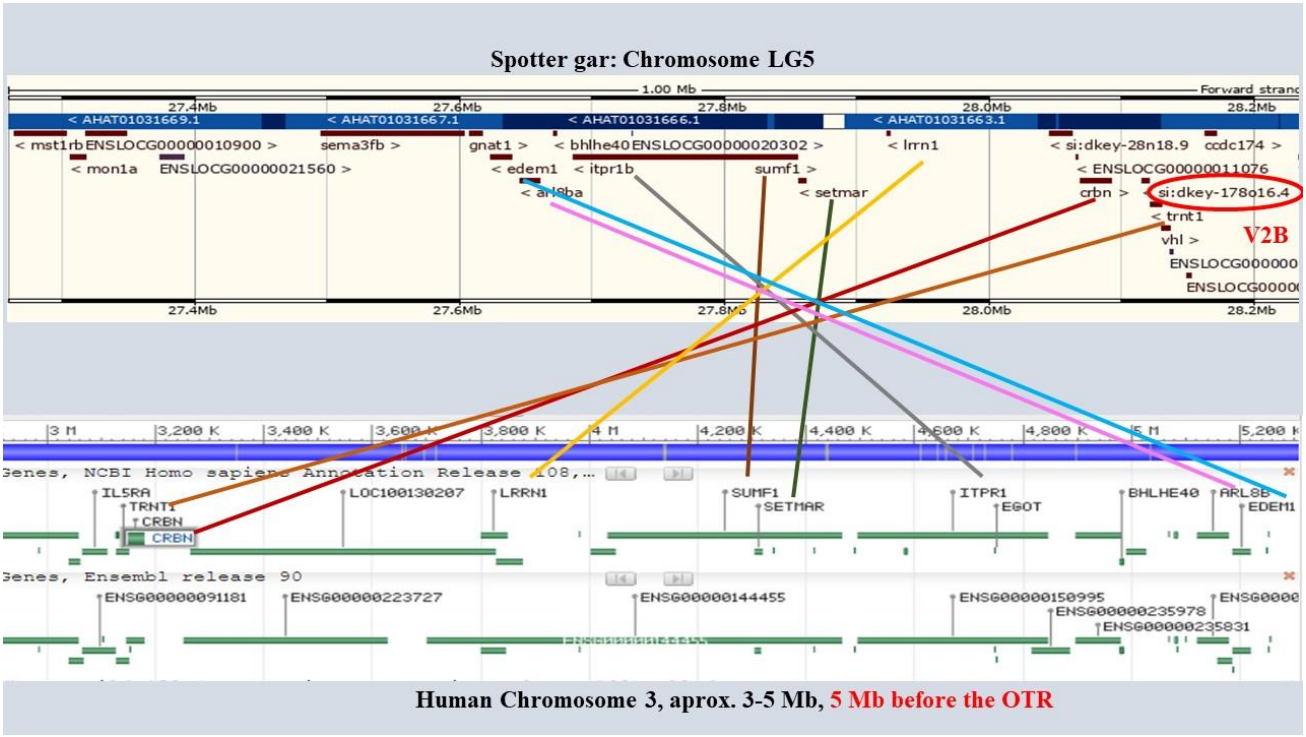


Figure S15: The genomic territory before the spotter gar VTR2B was also found in human chromosome 3 (3-5 Mb), 5 Mb before the location of human OTR.

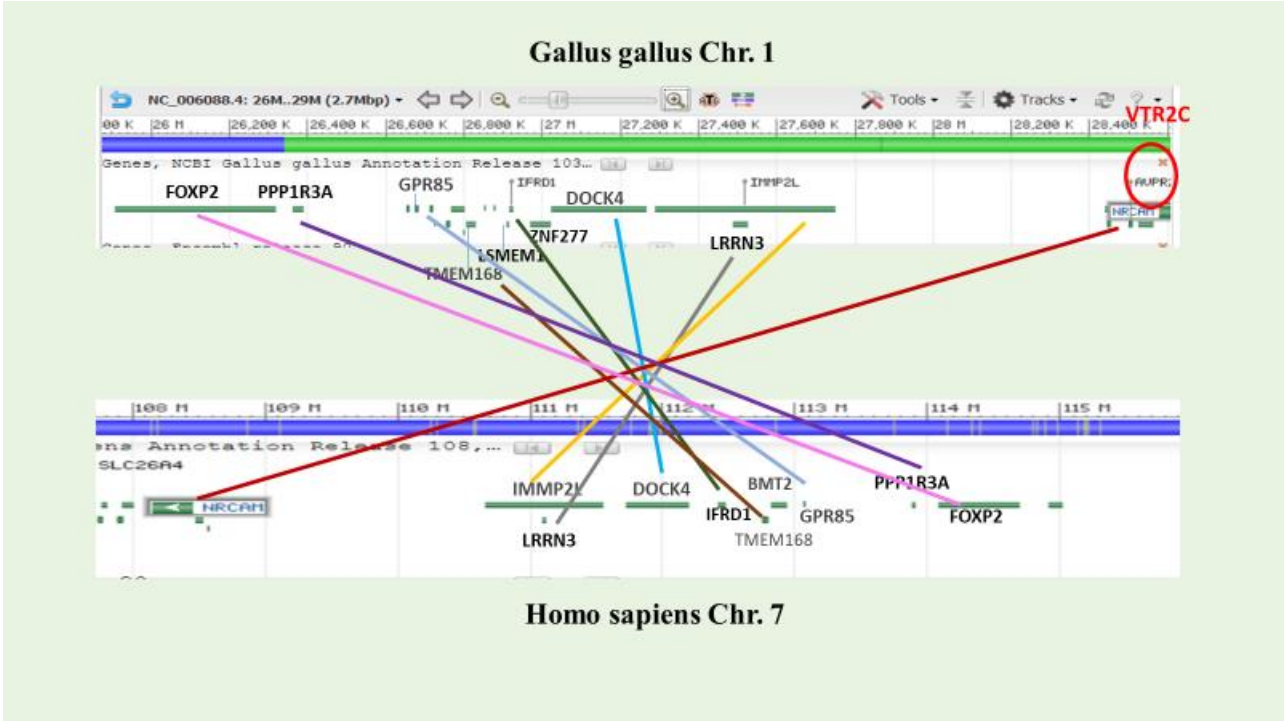


Figure S16: The genomic territory before the chicken VTR2C was found in human chromosome 7 (100-115 Mb).

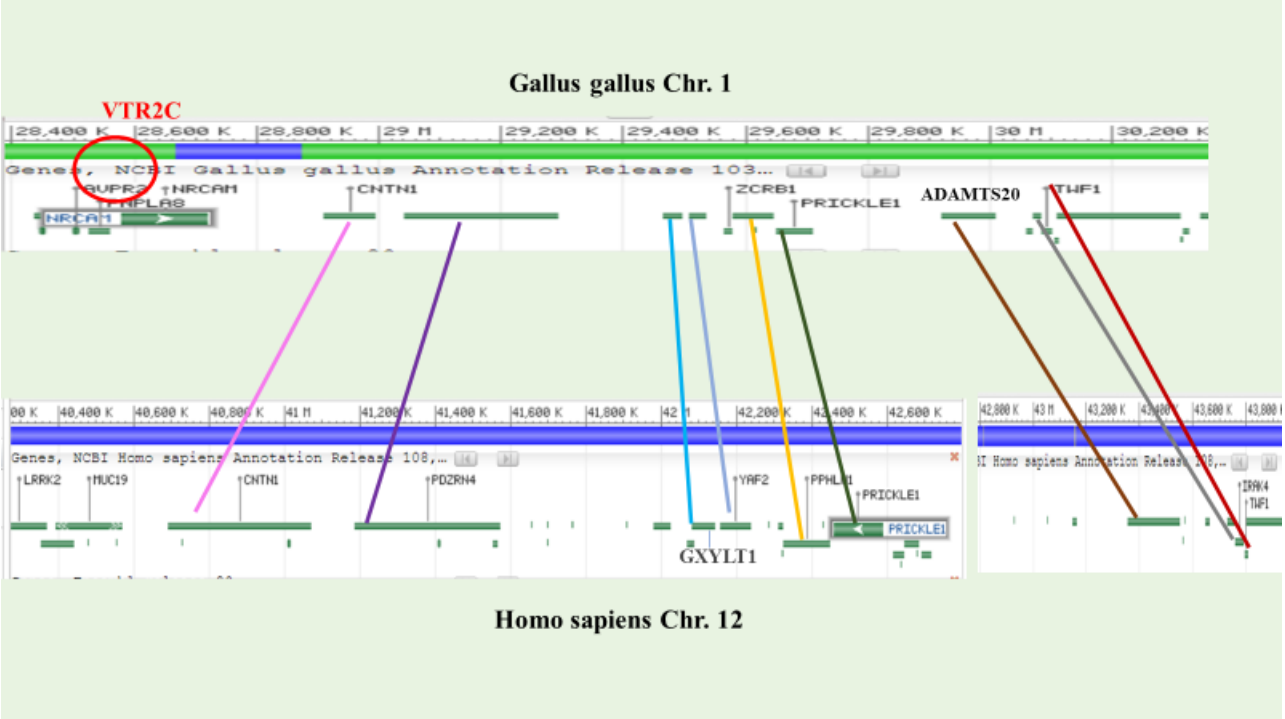


Figure S17: The genomic territory after the chicken VTR2C was found in human chromosome 12 (40-43 Mb).

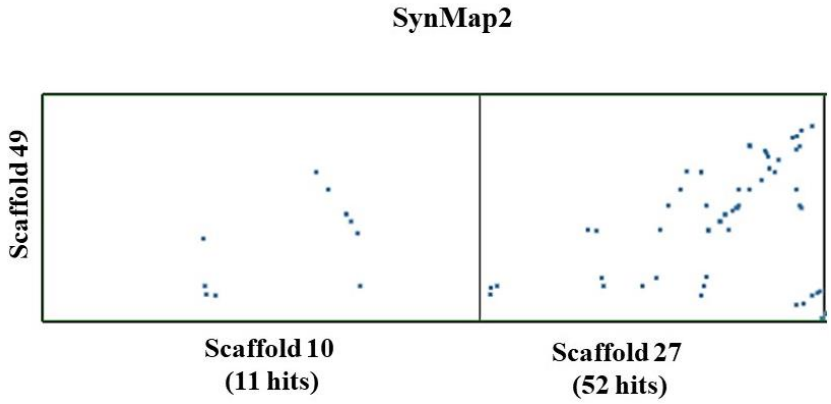


Figure S18: SynMap2 dotplot between sea lamprey scaffold 49 and scaffolds 10 and 27. (Parameters: Maximum distance between two matches:10 genes; Minimum number of aligned pairs: 3).