Supplementary Information

Structural Permissivity of the Oxytocin Receptor as a Driver of Oxytocin Pathway Evolvability in New World Monkeys

Bruna Oliveira Missaggia, Pedro Vargas-Pinilla, Tomas Marques-Bonet, Juliana Silva Bernardes, Márcio Dorn, and Maria Cátira Bortolini

Corresponding: Maria Cátira Bortolini
E-mail: maria.bortolini@ufrgs.br

Contents of this file

Supplementary Information

SI Text Section 1: Source and Data Acquisition.

1.1: Obtaining sequences and alignment. The annotated sequences of the *OXT* and *OXTR* genes were retrieved from public databases or obtained from the work of Vargas-Pinilla et al. (1). The databases used to recover sequences were Ensembl (http://www.ensembl.org/; (2)) and GenBank (https://www.ncbi.nlm.nih.gov/; (3)). For the unannotated sequences described in Kuderna et al. (4); Accession: PRJEB67744), we downloaded the complete genomes at the scaffold level. We performed exon BLAST using the canonical human sequence as a query and retained the result with the highest affinity for each species. For the OTR CDS, we merged the sequences of the two amino acid-coding exons. The annotated nucleotide sequences were aligned using Geneious R7.0.6 software (https://www.geneious.com) with the canonical *Homo sapiens* reference. Non-annotated sequences obtained via BLAST were added one by one to the multisequence alignment (MSA) and aligned using the MUSCLE algorithm (5), as implemented in the AliView v.1.28 software (6). The *OXTR* alignment was trimmed to remove the terminal stop codon (TGA), resulting in a final MSA CDS length of 1167 bp across 238 primate species and 274 samples. For *OXT*, sequences from 249 primate species and 290 samples were obtained. The sequences were translated into amino acids to assess the quality of the alignments. The quality of the amino acid alignment was assessed using BMGE (Block Mapping and Gathering with Entropy) (7), which identifies and removes low-quality columns based on entropy and gap content. A stricter evaluation was applied using the BLOSUM95 substitution matrix (-m BLOSUM95), and the quality was assessed on a position-by-position basis by setting the sliding window size to 1 (-w 1). Despite these stringent settings, no positions were removed, indicating that all columns in the alignment met the quality criteria. All other parameters were set to their default values. The sequence references for each species are listed in Table 1.

1.2: Taxonomy. The taxonomic classification adopted follows the most used system (8, 9), which divides the parvorder Platyrrhini into five families, as proposed by Rylands and Mittermeier (10). To maintain consistency across different data sources, whenever necessary we updated species nomenclature according to the IUCN (International Union for the Conservation of Nature). For example, several species formerly classified within the genus *Callicebus* (*Callicebus cupreus*, *Callicebus donacophilus*, etc.) were reclassified into the genus *Plecturocebus* (11). Some species of the genus *Saguinus* were reclassified into the genus *Leontocebus* (12). In the case of subspecies that were recently elevated to species status, such as *Perodicticus potto ibeanus* (now *Perodicticus ibeanus*, (13)), we retained all three names according to the data source, but they were treated as distinct species in analyses of intraspecific and interspecific variation

1.3: Phylogenetic information. The phylogenetic tree required for the PAML (14) and `phylolm` (15) analyses was constructed using the TimeTree website (https://timetree.org/; (16)). TimeTree is a public repository that compiles information on species divergence times, integrating molecular data currently covering 137,306 species. For each analysis, we input the corresponding species list and obtained a phylogenetic tree that includes them, along with their respective branch lengths. When necessary, modifications to the Newick files generated by TimeTree, such as unrooting the trees and marking the foreground, were performed using Python (17) scripts.

1.4: Phenotypic Data. In a broad sense, social monogamy can be defined as the tendency in some social species to form pairs that remain in close spatial proximity and may cooperate in raising offspring (18), often for successive breeding seasons (19). Social monogamy, therefore, is a complex phenotype encompassing several components: pair-living as a form of social organization, pair bonding as a social structure, a mating system that may involve both sexual and genetic monogamy, and cooperative or biparental care (20–22). There are some conceptual differences in the definition of this phenotype across various studies. Lukas and Clutton-Brock (20) classified species as socially monogamous if only one pair reproduces, regardless of whether other individuals are present in the group. This definition emphasizes the element of sexual monogamy (23). In contrast, Shultz et al. (24), Kappeler and Pozzi (25), and Olivier et al. (26) define social monogamy (also referred to by these authors as pair-living) as occurring when there are no other co-resident individuals. Olivier et al. (26) further describe pair-living as "repeated observations over several months of a uniquely identified adult male and an adult female, with or without dependent offspring, sharing a home range that overlaps substantially (>50%) but not with others (<50%). Pairs may forage together or separately." However, Shultz et al. (24) and Kappeler and Pozzi (25) consider species that forage alone as solitary, even if they share the same territory and sleeping sites. While Lukas and Clutton-Brock (20) classify each species into a single category, Shultz et al. (24) and Kappeler and Pozzi (25) allow for multiple classifications for the same species. Olivier et al. (26) also assign a primary form of social organization to each species but account for intraspecific variation in social organization in their research. Both conceptual differences and new studies may lead to species being classified differently across data compilations. For instance, fieldwork has shown that some nocturnal species in the suborder Strepsirrhini, often assumed to be solitary, actually live in pairs (26). French et al. (21) highlight that the diversity of components within social monogamy across species is so broad that they refer to it as "monogamy à la carte." These differences have led some authors (21–23, 25, 26) to advocate for individualized studies of the components of social monogamy or for contextualizing the concept by considering the different elements it comprises. Huck et al. (23) argue that the term monogamy should be reserved for reproductive and mating behaviors, and that social organization elements should be referred to as "pair-living" rather than social monogamy, as is often done. However, there is a lack of available data on the various components that make up social monogamy for a significant number of species, making comprehensive studies difficult. In any interspecies comparison, the quality of data is inevitably variable, as some species are more thoroughly studied than others. Additionally, field observations are challenging, requiring significant time and investment from researchers, leading to incomplete or biased data, particularly when a large sample size is crucial for comparative analyses. Despite these conceptual ambiguities, numerous comparative studies have been conducted using these data, yielding important inferences about the history and origin of the trait. One conclusion is that closely related species often share the same social or mating system (24, 25), suggesting the influence of genetic factors. Moreover, paternal care is frequently associated with both social and sexual monogamy and tends to evolve more easily in pair-living species (20, 27). Therefore, the characteristics that comprise what is commonly referred to as "social monogamy" are not independent of each other, and they may have common biological underpinnings that can be revealed through analyses aimed at identifying the genetic components behind these traits. In this study, to classify species in relation to the social monogamy phenotype (presence or absence of the trait), we adopt a broad definition, drawing primarily from references with a substantial number of primate (4, 20, 24–26). For studies that use more than one classification for a species, we classify it as socially monogamous if one of the classifications is "pair-living." In cases of classification discrepancies among studies, we consulted additional sources to support the adopted categorization. Table 2 presents phenotypes and corresponding references for each species. The dataset contains phenotypic information of 200 primate species.

SI Text Section 2: Evolution of OT and OTR variation.

2.1 Analysis of missense variation in primates. The analysis of amino acid variation at homologous positions was performed using Python (17) scripts with the Pandas (28) and NumPy (29) libraries. The MSA alignment of translated orthologous sequences was converted into a dataframe, enabling the identification of conserved positions across all species as well as sites with interspecific and intraspecific variation. Sites with ambiguous amino acids or gaps were excluded from the analysis to ensure that potential issues in data acquisition did not interfere with the results.

All amino acids in human oxytocin (Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly) participate in binding to OTR within the orthosteric ligand-binding pocket. The cyclic portion (residues 1–6) of oxytocin inserts deeply into the receptor's binding pocket, while the C-terminal tripeptide (residues 7–9) is oriented towards the extracellular loops (30, 31). To interpret the possible evolutionary implications of the variation observed in primates, we examined the specific function of each amino acid in its interaction with OTR. Table 3 shows which form of oxytocin is present in each primate sample from this study.

The OTR positions were analyzed concerning their location within the main functional domains of the protein and the information cited in the scientific literature regarding their functional significance in *Homo sapiens*. For this analysis, the protein was divided into three regions. The extracellular portion includes the three extracellular loops (ECL1–ECL3) and the N-terminal region, which contribute to ligand recognition and binding. The transmembrane region consists of the seven transmembrane helices (TM1–TM7), forming the receptor core responsible for ligand interaction. The intracellular portion, including the intracellular loops (ICL1–ICL3) and the C-terminal region, is involved in interactions with G proteins and other signaling molecules (32). Information on the positions corresponding to each domain was retrieved from the UniProt database (https://www.uniprot.org/; (33)).

For the quantitative analyses of the oxytocin receptor (OTR), scientific literature on oxytocin interaction sites consulted comprised Wesley et al. (34), Koehbach et al. (35), Waltenspühl et al. (30), Meyerowitz et al. (36) and Musiani et al. (31). The interaction between OTR and the antagonist retosiban was analyzed based on the structural data provided by Waltenspühl

et al. (32). Key positions with structural importance and other residues implicated in receptor activation, including disulfide bridges, cation-π-cation interactions, and various activation motifs, were extracted from Koehbach et al. (35), Waltenspühl et al. (32), Waltenspühl et al. (30), and Meyerowitz et al. (36). Positions associated with the G protein binding domain were retrieved from Gimpl and Fahrenholz (37), Waltenspühl et al. (30), and Meyerowitz et al. (36). Post-translational modification sites—including phosphorylation, ubiquitination, and glycosylation—were obtained from GPCRdb, as accessed on February 11, 2025 (38). From this literature, a set of 83 positions was compiled: 8, 15, 26, 32, 33, 34, 35, 36, 38, 39, 42, 46, 73, 74, 85, 88, 92, 93, 96, 99, 100, 103, 104, 112, 116, 119, 120, 123, 136, 137, 140, 141, 144, 171, 175, 187, 188, 189, 191, 200, 201, 204, 209, 223, 227, 230, 234, 235, 270, 272, 273, 284, 288, 291, 292, 295, 298, 299, 306, 307, 311, 312, 315, 316, 318, 319, 325, 326, 329, 335, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 366, 370. Additionally, to ensure a more precise analysis, a subset of the aforementioned positions was selected, encompassing only those critical for human OTR function and its interaction with oxytocin (OT) according to the most recent literature: 34, 42, 73, 96, 100, 103, 104, 112, 116, 119, 136, 137, 171, 187, 188, 200, 201, 204, 273, 284, 288, 291, 295, 316, 325, 326, 329 (30–32, 36).

Fig. 1 presents the proportional distribution of conserved, interspecific, and intraspecific variable sites across the three OTR domains (Extracellular, Transmembrane, and Intracellular). The bar chart categorizes these sites based on their citation in the literature, with darker shades representing positions identified as relevant for OTR function and interaction. The conserved positions, in which all primates in the dataset share the same amino acid, are likely crucial for the function and interaction of proteins in primates. Positions exhibiting intraspecific variation are presumably less relevant and are expected to be cited less frequently. Positions displaying only interspecific variation may be strong candidates for explaining differences between species, particularly if they are also cited in the literature as functionally relevant for *Homo sapiens*.

2.2 Evolutionary analyses PAML. To investigate evolutionary patterns in primate genes, we utilized the Phylogenetic Analysis by Maximum Likelihood (PAML v4.10.6 (14)) software, which employs maximum likelihood methods to evaluate molecular evolution and test models of positive selection against null models (without selection). The parameter ω ratio (d_N/d_S) quantifies selective pressure by comparing the rate of nonsynonymous substitutions (d_N), which alter amino acids and may be subject to selection, with the rate of synonymous substitutions (d_S), considered neutral. Values of $\omega < 1$ indicate purifying selection (elimination of deleterious mutations), $\omega = 1$ suggests neutrality, and $\omega > 1$ points to positive selection (accelerated fixation of advantageous mutations) (39).

It is important to highlight that ω represents the relative rate of selected versus neutral fixation events (39). Within a single species, strong positive selection is expected to lead to the fixation of advantageous sites. This implies that sites exhibiting both intraspecifchiic variation and $\omega > 1$ should not be interpreted as having undergone positive selection (40). Instead, such sites are more likely indicative of ancestral polymorphisms. Due to this, we ran the program in two ways. In one approach, we included only one sample per species and excluded all sites with intraspecific variation (marking them as 'N' and using the parameter `cleandata=1`). Thus, this analysis focused exclusively on sites fixed within species. In a second approach, we retained intraspecific variation. In cases of sequence differences between individuals of the same species, we inserted them into the tree and retained both differing sequences in the FASTA file. If positive selection is detected at sites with intraspecific variation, they will be considered false positives.

Using the CODEML package, we estimated the ω ratio across codons (Site models; NSsites test) and across codons and branches within the phylogeny (Branch-site models). The complete coding sequences of the *OXT* (224 species) and *OXTR* (204 species) genes were analyzed, considering only species with available genetic and phylogenetic data (see Table 4). The phylogeny was constructed as described in Section 1.3. The analyses were conducted using unrooted trees, since the substitution model is time-reversible and assumes that the branches near the root evolve under the same process (41).

2.2.1 Site models. Site models assume that codons experience distinct selective pressures, leading to variation in ω. Positive selection is characterized by codons with $\omega > 1$. Likelihood ratio tests (LRTs) were employed to compare null models (M1a, M7, and M8a) that constrain $\omega \leq 1$ with alternative models (M2a and M8) that allow for $\omega > 1$. The LRT statistic, or twice the log-likelihood difference between the two compared models ($2\Delta\ell$), was used in a chi-square test with the degree of freedom to be the difference in the number of free parameters between the two models. The LRT statistic and chi-square test were conducted using Python (17) with the Pandas (28) and SciPy (42).

We tested for positive selection by comparing three model pairs using CODEML: (1) M1a (Nearly Neutral) vs. M2a (Positive Selection), df = 2 (43, 44), (2) M7 (beta) vs. M8 (beta&ω), df = 2 (39), and (3) (beta&$\omega_s = 1$) vs. M8 (beta&ω), df = 1 (43, 45). When the models M2a or M8 provided a significantly better fit to the data, it suggested that a fraction of codons evolved under positive selection (46). Bayes Empirical Bayes (BEB) analysis, implemented in CODEML (PAML v4.10.6), was used to identify the codon sites most likely to be under positive selection, providing site-specific probabilities and ω estimates (44).

For *OXT*, when including intraspecific variation, null models (M1a, M7, M8a; $\omega \leq 1$) outperformed positive selection models (M2a, M8) in all comparisons (Table 5). Under M8a, 99.999% of *OXT* sites were under purifying selection, with ω values following a beta distribution ($p = 0.03250$, $q = 0.33376$) (Table 6). In the analysis excluding intraspecific variation (codon 2 removed), the M7 vs. M8 comparison showed marginal significance ($2\Delta\ell = 6.83$, df = 2, $p = 0.033$). However, M8 assigned 99.999% of sites to a beta distribution ($p = 0.01359$, $q = 0.08949$; $\omega < 1$), with a negligible proportion (0.00001%) having $\omega = 3.94446$ (Table 7 and Table 8). Bayes Empirical Bayes (BEB) analysis identified no sites under significant positive selection (posterior probability > 0.95); the highest probability (0.747) corresponded to the eighth position of the nonapeptide, with

$\omega = 0.747 \pm 0.379$. Other model comparisons were non-significant; therefore, there is no robust evidence of positive selection in *OXT*.

For *OXTR*, when considering intraspecific variation, models allowing positive selection (M2a, M8) fit better than null models. M8 significantly outperformed M8a ($2\Delta\ell = 5.70, df = 1, p = 0.0169$), assigning 99.13% of sites to a beta distribution ($p = 0.18340$, $q = 2.40521$; $\omega < 1$) and 0.87% to a class with $\omega = 2.19070$ (Table 9 and Table 10). BEB identified two sites (6A, 47C; posterior probability = 0.999) and one marginal candidate (375R; probability = 0.949). Positions 6 and 47 were previously proposed as potential targets of positive selection (1, 47). However, these three sites exhibited intraspecific variation across multiple species, a pattern more consistent with ancestral polymorphisms than with adaptive evolution (40). Our data reveal intraspecific variation at position 6 in two species (*Mico humeralifer* and *Cacajao melanocephalus*), position 47 in four species (*Ateles belzebuth*, *Ateles geoffroyi*, *Cacajao melanocephalus*, *Plecturocebus moloch*), and position 375 in five species (*Alouatta caraya*, *Aotus azarae*, *Ateles geoffroyi*, *Lagothrix lagotricha*, *Plecturocebus cupreus*). This variation, or part of it, may reflect errors in sequencing or other steps in data acquisition (48), and thus requires careful interpretation before dismissing the possibility of relevance. We verified that none of these positions overlapped with residues cited as important for OTR function, as would be expected for sites that have undergone positive selection (49), nor did they exhibit variation associated with the investigated phenotype. This reinforces the likelihood that these sites are not under positive selection. When excluding variable sites, all likelihood ratio tests were non-significant. The null model M8a best described the data, with 99.999% of sites under purifying selection (beta distribution: $p = 0.16774$, $q = 2.31793$; $\omega < 1$) and a negligible fraction (0.00001%) evolving neutrally ($\omega = 1$). The overall d_N/d_S ratio (0.0625) confirmed strong purifying selection, with no evidence of adaptive evolution (Table 11 and Table 12).

These analyses indicate that both *OXT* and *OXTR* are predominantly under purifying selection, with no robust evidence of positive selection at specific codon sites when intraspecific variation is accounted for.

2.2.2 Branch-site model. To analyze the presence of positive selection in Platyrrhini branches, we employed the modified branch-site model A (44, 50) implemented in CODEML (PAML v4.10.6). This model allows the ω ratio (d_N/d_S) to vary both among protein sites and across branches in the phylogenetic tree. The test evaluates the hypothesis that subsets of amino acid sites in specific branches (defined a priori) are subject to positive selection ($\omega > 1$). Platyrrhini branches were designated as foreground branches, while all other lineages were treated as background. The alternative model allowed $\omega > 1$ in the foreground branches, whereas the null model fixed $\omega = 1$ in these branches. Statistical significance was assessed using a Likelihood Ratio Test (LRT), comparing the likelihoods of the models and considering a null distribution comprising a 50:50 mixture of point mass 0 and χ^2_1. Consequently, the p-value was halved (14).

For the *OXT* gene, branch-site model tests detected no evidence of episodic positive selection in Platyrrhini branches. When including all codons, the likelihood ratio test (LRT) was non-significant (LRT = 0, p = 1.0; Table 13). Although codon 2 exhibited a high Bayes Empirical Bayes (BEB) posterior probability (Prob = 0.967), the estimated ω_2 remained 1.0, contradicting positive selection. This site also showed intraspecific variation, suggesting any inferred selection would be a false positive. Excluding codon 2, the analysis similarly yielded non-significant results (LRT = 0, $p = 1.0$), with no sites under positive selection (BEB > 0.95) and $\omega_2 = 1.0$ in foreground branches (Table 14).

For the *OXTR* gene, analyses including intraspecific variation revealed no support for positive selection in Platyrrhini branches (LRT = 0, $p = 1.0$; Table 15). Log-likelihood values for alternative and null models were nearly identical (lnL = -11074.947806 vs. -11074.947891), with $\omega_2 = 1.00000$ in foreground branches. Excluding variable sites, results remained non-significant (LRT = 0, $p = 1.0$), as model likelihoods did not differ meaningfully (lnL = -4548.100554 vs. -4548.353707) and $\omega_2 = 1.00000$ was retained (Table 16).

These findings indicate no significant episodic positive selection in Platyrrhini lineages for either *OXT* or *OXTR* genes under the tested conditions, whether intraspecific variation was included or excluded.

SI Text Section 3: Missense Variants in OT/OTR and Social Monogamy.

3.1. Encoding Genetic Variants into Features. The multisequence alignment (MSA) files in FASTA format, containing the amino acid sequences of OT and OTR, were processed to create dataframes. In these dataframes, each row corresponds to a sample, while each column represents a position in the sequence. To ensure clarity and consistent identification, the positional columns were labeled with the prefix 'O' or 'R', indicating OT and OTR, respectively, followed by the numerical position.

Positions with no variation among samples or containing missing data were excluded from the analysis. Additionally, positions with intraspecific variation were removed, as the study focuses exclusively on interspecific phenotypic differences. This filtering resulted in 2 positions with interspecific variation in OT and 92 in OTR. Subsequently, the dataframes for OT and OTR were merged, retaining only the 212 primate species common to both datasets.

The genetic variations were encoded as features using the dataframe containing 94 positions with interspecific variation across 200 species for which data on the social monogamy phenotype were available. Dummy encoding was applied, transforming the variants into binary (dichotomous) features that could take values of 0 or 1. To reduce redundancy while preserving positional information, positively and negatively correlated variants were combined and annotated using the symbols '+' and '-', respectively.

For example, a combined feature such as 'R66_T+R343_R-R66_I+R343_H' represents specific combinations of amino acid variants at positions 66 and 343 of the oxytocin receptor. Binary values indicate the presence or absence of these specific combinations. For instance, species with a value of 1 for this feature possess amino acid T at position R66 and R at position

R343, while lacking I at position R66 and H at position R343. Conversely, species with a value of 0 lack T at position R66 and R at position R343 but have I at position R66 and H at position R343.

These procedures were performed in Python (17) using the following packages: Pandas (28), NumPy (29) and scikit-learn (51). As a result of this process, a dataframe with 143 features was generated for the 200 species with available data on the social monogamy phenotype. None of the features consists of variations in both genes; 138 features (96.5%) represent variations in the OTR, while 5 features (3.5%) represent the variation found in OT. Most features, 126 (88.11%), represent genetic variation at a single position, whereas 17 features (11.8%) represent genetic variation combining more than one position.

3.2. Evaluation of Features' Explanatory Capacity for Phenotype. The potential explanatory capacity of the features regarding the phenotype was evaluated using a classification decision tree, which incorporated all phenotypic outcome information and features. The algorithm utilized information on the class membership of instances (species), categorized as either exhibiting or not exhibiting social monogamy, to generate decision rules based on the features. This process resulted in a hierarchical structure composed of nodes, branches, and leaves. Nodes contain the features used to separate species, and branches direct species to subsequent nodes or leaves based on a given decision rule. The goal of a classification decision tree is to separate instances of different categories using the most relevant features. Leaves containing instances from only one class are considered pure, while leaves containing instances from multiple classes are referred to as impure.

Species located in impure leaves indicate that the features representing candidate genetic variations were insufficient to distinguish them. It is assumed that the coding variation in these genes is not associated with the phenotype in these species, and they are excluded from subsequent analytical steps.

The decision tree for this analysis was generated using the `DecisionTreeClassifier` implementation from the scikit-learn library (51), which is based on the CART algorithm (52). At each node, the algorithm evaluates which feature best separates the instances by minimizing category mixing (impurity). The criterion used to select the feature is Gini impurity, a value ranging from 0 to 0.5, representing the proportion of the minority category in each node. A Gini value of 0.5 indicates an equal proportion of instances in each category, while a value of 0 indicates that all samples belong to the same category. The feature selected at each node is the one that most effectively divides the data into correct classes, resulting in the lowest possible Gini value.

The analysis was conducted using a dataframe containing 200 primate species (see Table 4) and 143 features. Python (17) was used for data manipulation and analysis, employing libraries such as Pandas (28) and NumPy (29) for preprocessing. The decision tree correctly classified 187 of the 200 species into pure leaves as either monogamous or non-monogamous but failed to distinguish 13 species belonging to the Strepsirrhini infraorder, specifically the Lemuriformes families Indriidae and Lemuridae. The monogamous *Indri indri* and *Eulemur rubriventer* shared identical feature profiles with nine non-monogamous species: *Propithecus diadema*, *P. perrieri*, *Lemur catta*, *Eulemur rufus*, *E. albifrons*, *E. sanfordi*, *E. flavifrons*, *E. macaco*, and *E. coronatus*. Likewise, the monogamous *Hapalemur occidentalis* was indistinguishable from the non-monogamous *Prolemur simus*. These 13 species were excluded, leaving a dataset with 187 species and the 142 features of the candidate genes that vary among them for subsequent analytical steps.

3.3 Identifying Broad Set and Testing Feature Associations. To investigate genetic features potentially associated with social monogamy in primates, we employed multiple complementary and independent approaches, each testing a distinct hypothesis. These procedures are herein referred to by the following abbreviated names for clarity: BroadSet_RF, Fisher_Exact, PhyloLogReg, and RF_PermImp. All these procedures were performed with the 187 species listed in the Broad_Set_Assoc column of Table 4.

3.3.1 Identification of the Broad Set of Important Features (BroadSet_RF). The broad set of features associated with the phenotype (BroadSet_RF) is defined as the minimum set of important features required to correctly classify all species. To correctly classify all 187 species, we performed a preliminary evaluation of the relative importance of 142 genetic features that varied among these species. The importance of each feature was quantified using the `RandomForestClassifier` from the scikit-learn library (51), an ensemble machine learning algorithm that constructs multiple decision trees on different subsets of data.

To estimate feature importance, 1,000 decision trees were generated using different random states. Each tree utilized a subset of features equivalent to the square root of the total number of features and included all species in the dataset. Feature importance within a single tree was determined by the reduction in impurity caused by the feature at the tree's nodes. The overall importance of each feature was calculated as the average importance score across all trees. This procedure followed the default method of Random Forest (.feature_importances_), with the sole modification of increasing the number of estimators (n_estimators) from 100 to 1,000. The relative importance of each feature is shown in Dataset S1.

The features were ranked by importance, and the minimum set of features capable of correctly classifying all species was selected. This was accomplished by incrementally training decision trees: starting with the most important feature identified by Random Forest, additional features were progressively added until all species were correctly classified.

While the use of 1,000 decision trees may lead to overfitting, this approach was intentionally employed to align with the exploratory objective. Overfitting facilitates the identification of all potentially relevant features, even those that are redundant or specific to a small number of species. This inclusive strategy ensures the retention of genetic variations that might otherwise be excluded in models focused solely on generalization. Importantly, the stopping criterion—ensuring all species are correctly

classified—guarantees that even variations associated with the phenotype in one or a few species are considered into subsequent analyses.

By relying solely on generalization-based methods, specific variations with potential biological relevance could be inadvertently eliminated. In contrast, this broader approach prioritizes inclusivity, reflecting the complexity of biological systems. Previous studies have shown that different sets of features—often with little to no overlap—can achieve similar predictive performance (53–55). This phenomenon arises because many genetic variables are correlated, and only a subset may be essential (56). Retaining all potentially relevant features allows for the detection of complex interactions and subtle patterns that are crucial in understanding phenotypic distinctions.

It is important to emphasize that selecting a feature based on this criterion does not necessarily imply that the genetic variation it represents is directly associated with the phenotype. Instead, the process retained all features potentially relevant to distinguishing phenotypic classes, including those with individual significance and those that may interact with others. By preserving a comprehensive set of features, this approach enhances result interpretability and enables the detection of nuanced biological interactions that might otherwise be overlooked in models solely focused on predictive accuracy. Genetic variations represented by these features are interpreted alongside additional evidence, such as signals of positive selection, coevolution, functional predictions, and statistical association methods.

All analyses were conducted in Python (17) using the following packages: Pandas (28), NumPy (29), scikit-learn (51), matplotlib (57), and seaborn (58). A total of 39 features were required to correctly classify the 187 species with respect to the social monogamy phenotype (Table 17). The names of these 39 features are shown in Table 17 and Fig. 2.

3.3.2 Association of Features and Phenotype Using Fisher's Exact Test (Fisher_Exact). Fisher's exact test (59) was applied individually to each of the 142 genetic features in a dataset of 187 primate species. The test is particularly useful for small sample sizes or sparse data, it assessed whether the presence of a genetic variant was associated with the social monogamy phenotype.

For each feature, a 2x2 contingency table was constructed. One axis represented the presence or absence of the genetic variant, and the other axis indicated whether the species exhibited social monogamy or not. The test calculates the exact probability of observing the given data distribution, assuming the null hypothesis of no association between the genetic feature and the phenotype. To control multiple comparisons, Bonferroni correction (60) was applied by dividing the chosen type I error probability ($\alpha = 0.05$) by the number of independent hypotheses tested. This procedure is referred to as 'Fisher_Exact' in figures and tables.

The analysis was conducted using Python (17) with Pandas (28), NumPy (29), scikit-learn (51), SciPy (42), and Statsmodels (61). Dataset S2 presents the raw and Bonferroni-corrected p-values, while Fig. 2 and Table 18 highlight the 17 features with significant p-value ($p < 0.05$) after correction.

3.3.3 Association of features and phenotype using phylogenetic logistic regression (PhyloLogReg). The phenotypic trait analyzed is a binary dependent variable, representing a trait with two discrete states (presence or absence of the behavior). For this reason, phylogenetic logistic regression (62) was employed. This method evaluates the effects of independent variables (features) on the phenotype (binary dependent variable) as it evolves along a phylogeny. The approach was used because species share evolutionary history, violating the assumption of error independence required by traditional statistical methods (63). Non-independent residuals, common among closely related species, can result in Type I errors, with artificially low p-values leading to the rejection of true null hypotheses (64–67).

Methods incorporating phylogenetic information have been developed to address these limitations, estimating correlations in comparative data across multiple species while accounting for shared evolutionary history through phylogenetic trees (62–64, 68, 69). However, when a phenotypic variation is exclusive to a single lineage, these methods may remove the signal due to their focus on patterns of evolutionary convergence or parallel fixation of ancestral polymorphisms. This can result in false negatives, where variations specific to taxa are treated as part of the phylogenetic component shared by common ancestry.

The analysis was performed in R environment (70) using the `phyloglm` function from the `phylolm` package (version 2.6.2) (15). This package depends on the ape package, for which version 5.0 was used (71). The required unrooted phylogenetic tree was generated as explained in Section 1.3. In `phyloglm` function, phylogenetic signal is controlled by the parameter α, which ranges from 0 to 1. A value close to 0 indicates strong phylogenetic dependence (closely related species exhibit more similar traits), whereas a value close to 1 suggests an absence of phylogenetic signal. The α parameter is iteratively estimated until it reflects the observed phylogenetic signal in the sample data. Subsequently, the values of β_0, β_1, and the p-value are obtained. We named the procedure developed to apply the `phyloglm` function individually to each feature, store the results and correct for multiple tests by Bonferroni 'PhyloLogReg'.

The α values, β_1 (the coefficient quantifying the effect of the feature on the behavior), p-values, and Bonferroni-adjusted p-values for each feature are presented in Dataset S3. The top 10 features ranked in the PhyloLogReg are shown in Table 19. The α values for all 142 features were close to zero, indicating strong phylogenetic signal, and no p-value below 0.05 was identified.

3.3.4 Association of features and phenotype using Random Forest p-value (RF_PermImp). The Fisher's exact test and phylogenetic logistic regression are easily interpretable methods; however, they can only identify features that provide a significant amount of information when analyzed in isolation from other features. Since biological processes are expected to involve complex interactions between variables, methods that evaluate the relationship of each feature with the phenotype independently of

others may fail to identify some informative features. Variable importance measures (VIMs), such as the '.feature_importances_' provided by Random Forest, have the potential to highlight interacting effects. However, since they do not provide p-values by default, they are not statistically interpretable (72).

Although p-values for assessing the significance of Random Forest importance measures are not automatically generated, they can be derived using null distributions. An approach involves permuting the response variable, fitting the model, and calculating the importance of each feature during each iteration. The p-value is then determined as the proportion of iterations in which the permuted importance values are greater than or equal to the observed importance. By permuting only the phenotypes, which represent the target variable, all relationships between the features and the target are eliminated, thereby representing the null hypothesis of no association. Importantly, the natural relationships among genetic variations (features) remain intact. This strategy offers advantages over alternatives that permute features independently, such as the 'permutation importance' function in the scikit-learn 'inspection' module (51), which is computationally intensive and disrupts associations among features. Preserving the natural relationships among features ensures a more realistic null model, which better reflects the interactive effects observed in real biological data (73).

We implemented this procedure in Python (17), utilizing the `RandomForestClassifier` from the scikit-learn library (51) with default parameters. Null importance distributions were generated through 100,000 permutations of the target variable (phenotype). This method, here called 'RF_PermImp', was applied to the dataset comprising 187 species correctly classified and 142 features that exhibited variation among them. Additional libraries used in the analysis included Pandas (28), NumPy (29), and Statsmodels (61). Dataset S4 provides the p-values for each feature, both before and after Bonferroni correction (60). Fig. 2 and Table 20 highlight the 11 features with p-values below 0.05 following Bonferroni correction.

3.4 Functional Impact Assessment of Amino Acid Substitutions. To investigate the potential functional relevance of amino acid substitutions represented by the selected features, we employed two general and complementary classes of structure- and function-based approaches to prioritize variants for interpretation.

First, we considered the functional importance of specific protein positions based on previously published experimental data. The identification of residues with known functional roles in human oxytocin and its receptor OTR was conducted through a literature review. This analysis was based on the following references: (30–32, 34–38), as detailed in Section 2.1. These studies provided information on structural domains, interaction sites, and post-translational modifications, supporting the interpretation of amino acid substitutions in the context of protein conformation and molecular function.

Second, we employed two complementary computational strategies to evaluate the potential impact of the amino acid substitutions: (1) an analysis based on molecular characteristics, using the Grantham score (74); and (2) a computational pathogenicity prediction analysis, using the algorithms SIFT (75), PolyPhen-2 (76), and AlphaMissense (77).

Functional analyses are grounded in the assumption of homology and evolutionary conservation across species. They utilize available data for *Homo sapiens* as a reference to interpret the potential implications of selected amino acid variations, which are hypothesized to contribute to observed phenotypic divergence.

3.4.1 Granthan score: Molecular Characteristics of Amino Acid Substitutions. The Grantham score (74) is a widely used metric that quantifies the physicochemical dissimilarity between amino acids based on three key properties: side chain composition, polarity, and molecular volume. Each substitution is evaluated in a three-dimensional physicochemical space, where the Grantham score corresponds to the Euclidean distance between amino acids after applying weights to each property. These weights were originally calibrated using amino acid substitution rates estimated by McLachlan (78). Higher scores indicate more radical substitutions, reflecting greater divergence in physicochemical characteristics.

To aid in interpretation, Grantham scores have been categorized by Li et al. (79) as follows: conservative (0–50), moderately conservative (51–100), moderately radical (101–150), and radical (>150). This classification provides a standardized framework for assessing the potential structural and functional consequences of amino acid changes.

For each position with selected features, the Grantham score is calculated from the comparison between the amino acid of each species in the target class (i.e., species exhibiting social monogamy), with the most frequent amino acid in the non-target class (encoded as phenotype 0). When the most frequent amino acid is shared between both classes, the score is based on the second most frequent amino acid in the non-target class. For subsequent analyses, only radical substitutions observed in correctly classified socially monogamous species were considered.

3.4.2 Pathogenicity Prediction Analyses. In parallel, we conducted pathogenicity prediction analyses for missense variants corresponding to the identified features, using three widely adopted predictors: SIFT (75), PolyPhen-2 (76), and AlphaMissense (77).

The SIFT (Sorting Intolerant From Tolerant) algorithm predicts the potential functional impact of amino acid substitutions by analyzing the evolutionary conservation within protein families. It operates under the principle that positions highly conserved throughout evolution are likely critical for protein function, and thus substitutions at these sites are more likely to be deleterious. SIFT uses multiple sequence alignments obtained through PSI-BLAST to assess the degree of conservation and calculates the probability of tolerance for each possible amino acid substitution at a given position. The resulting score, known as the SIFT score, ranges from 0 to 1, representing the normalized probability that a substitution is tolerated. A substitution is predicted to be "deleterious" if the SIFT score is less than 0.05, whereas scores equal to or greater than 0.05 indicate a "tolerated" substitution (75).

PolyPhen-2 (Polymorphism Phenotyping v2) predicts the potential functional impact of amino acid substitutions using a combination of sequence- and structure-based features. The algorithm evaluates how a given mutation affects protein function by considering multiple predictive parameters, including the position of the substitution in conserved sequence alignments, the physicochemical differences between wild-type and mutant residues, and structural attributes such as accessible surface area and the presence within functional domains. These features are integrated using a Naive Bayes classifier, trained on datasets of known benign and damaging mutations (HumDiv and HumVar). PolyPhen-2 assigns a probabilistic score indicating the likelihood that a mutation is damaging, which is further interpreted qualitatively. Scores range from 0.0 to 1.0, with higher values indicating a greater probability of being damaging. The tool defines thresholds to classify mutations as "benign" (score close to 0), "possibly damaging" (intermediate scores), or "probably damaging" (score close to 1.0). Additionally, predictions may be labeled as “unknown” when data are insufficient for classification (76).

AlphaMissense predicts the functional consequences of amino acid substitutions by integrating three main strategies: (i) protein language modeling to learn context-dependent amino acid distributions; (ii) structural context derived from AlphaFold to understand spatial constraints; and (iii) fine-tuning on weakly labeled data based on human and primate population variant frequencies, avoiding biases from manually curated clinical databases. The model assigns a pathogenicity score to each variant, ranging from 0 to 1, where values near 0 indicate high confidence of being benign and values near 1 suggest pathogenicity. These scores are then calibrated using ClinVar data to reflect the probability of a variant being pathogenic. Based on calibrated thresholds that yield 90% precision, variants are classified into three categories: “likely benign” (low scores), “likely pathogenic” (high scores), and “uncertain” (intermediate scores between thresholds) (77).

SIFT and PolyPhen-2 scores were generated using the Ensembl Variant Effect Predictor (VEP; https://www.ensembl.org) with *Homo sapiens* as the selected species and HGVS notation of missense variants as input (transcript ID ENST00000217386.2). The 27 base pairs (bp) corresponding to the oxytocin nonapeptide are located at positions 3,071,713 to 3,071,739 on chromosome 20 (GRCh38/hg38 reference genome). These coordinates span amino acids 20–28 of Oxytocin-neurophysin 1, which were mapped to positions 1–9 of the oxytocin hormone to align with the reference numbering system adopted in this study. For the oxytocin receptor (transcript ID ENST00000316793.8), features positional annotations align with the *Homo sapiens* reference sequence, thus requiring no conversion for this protein. AlphaMissense predictions were obtained from the AlphaMissense platform (https://alphamissense.hegelab.org/) using UniProt accession codes (https://www.uniprot.org/; (33)) for *Homo sapiens* proteins (P01178 [Oxytocin-neurophysin 1] and P30559 [Oxytocin receptor]), retaining only positions selected.

These predictors facilitate the prioritization of variants for further interpretation. Variants classified with high pathogenicity scores by two or more algorithms are retained for subsequent analyses. Predictions for the genetic variants selected by features are provided in Table 21.

3.5 Interpretative Framework for Prioritizing Genetic Variants Based on Statistical, Evolutionary, and Functional Evidence. To systematically evaluate the relevance of genetic variants identified as associated with social monogamy by each approach, we developed an interpretative framework (see explanatory summary below). This matrix integrates signals of positive selection (assessed via site and branch-site models) and functional relevance inferred from two complementary sources: positions previously reported in the literature as functionally important for protein structure or interaction, and computational predictions of functional impact based on physicochemical divergence (Grantham score) and pathogenicity estimates (SIFT, PolyPhen-2, and AlphaMissense). This explanatory resource synthesizes the implications of variant-trait associations across methods and provides a structured rationale for prioritizing amino acid substitutions for the next stage of investigation. It also facilitates the interpretation of the results obtained through the different analytical approaches employed in this study.

Test	Implications of Selected Features	Positive Selection: Site Models	Positive Selection: Branch-site Model (Platyrrhini)	Functionally Relevant Site (Literature)	Putative Functional Impact (Grantham & Pathogenicity Predictions)
BroadSet_RF	Unique features may reflect weak non-linear associations, species-specific effects, or variants impacting few species. Without additional evidence, correlations are likely spurious.	Positive selection on exclusive variants suggests selective pressure linked to a related phenotype, not monogamy. Strengthens evidence of functional impact.	Similar to site models but specific to Platyrrhini. Indicates selection on a related phenotype influencing behavior in carrier species.	Variants at functional sites may modulate behavior. Combination with positive selection reinforces relevance. Case-by-case validation is required.	High Grantham scores (radical) or pathogenic predictions (SIFT/PolyPhen-2/AlphaMissense) may indicate functional impact on sites cited in the literature as relevant. Combination with positive selection reinforces relevance. Case-by-case validation is required.

Fisher_Exact	Associations not replicated in PhyloLogReg are likely taxon-specific or spurious due to phylogenetic confounding. Requires confirmation via selection or function.	Positive selection implies adaptation in primates associated with monogamy. Functional investigation is recommended.	Selection in Platyrrhini suggests localized adaptation. Investigate role in monogamy and molecular interactions.	Association at functional sites, even without selection, warrants investigation. Positive selection combined with functional relevance supports causal role.	Suggests adaptive significance if aligned with other evidence.
PhyloLogReg	Phylogeny-independent associations suggest convergent evolution or parallel fixation. Prioritize investigation even without additional evidence.	Positive selection supports generalized adaptation in primates. Variants likely directly influence the phenotype.	Selection in Platyrrhini supports specific adaptation. Combined with phylogenetic association indicates convergent evolution.	Functional sites linked to associations suggest convergent molecular mechanisms. Positive selection strengthens causality.	High predicted impact supports convergent evolution at functionally critical residues. Reinforces the causal interpretation of the variant.
RF_PermImp	Unique features indicate non-additive interactions. Require functional or selection support to validate relevance.	Positive selection implies primate adaptation, potentially mediated by genetic interactions.	Selection in Platyrrhini suggests adaptive interactions. Differences from Fisher_Exact indicate non-additive effects.	Non-additive interactions warrant structural or experimental analysis. Association at functional sites, combined with selection, supports causal role.	If combined with other evidence, it supports mechanistic role in non-additive interactions.

3.6 Analysis of Combinations of Variations Selected from the Interpretative Framework. Amino acid variations with plausible involvement in phenotypic differentiation, based on the interpretative framework, were retained for further analysis. Only groups of species for which the selected variations were informative in distinguishing the monogamy phenotype were included in this analysis; all others were excluded. A dataframe was constructed containing the species in which each variation occurred, grouped according to their respective phenotypic categories (socially monogamous or non-monogamous). Subsequently, combinations of amino acid residues were identified across the sample. Species were grouped by their specific amino acid combination, and the distribution of phenotypic categories within each combination was assessed.

A summary table was generated, reporting the number of species carrying each unique combination, along with their corresponding taxonomic families. Additionally, a social monogamy index was calculated for each combination by dividing the number of monogamous species from the total number of species carrying the combination. An index value of 0 indicates that none of the species with that combination are socially monogamous, a value of 1 indicates that all species are socially monogamous, and intermediate values reflect the degree of mixture between phenotypic categories within the combination.

To assess whether specific combinations are associated with social monogamy, we performed Fisher's exact tests for independence (59) between each unique combination and the monogamy phenotype. The odds ratio (OR) was computed for each table to measure the strength and direction of the association between the presence of the combination and the monogamous phenotype. An OR > 1 suggests a positive association (i.e., the combination is more frequent in monogamous species), while OR < 1 suggests a negative association. p-values were calculated using the Fisher's exact test (two-tailed). To account for multiple testing (one test per unique combination), Bonferroni correction (60) was applied to control type I error due to multiple comparisons. The analysis was conducted using Python (17) with Pandas (28), NumPy (29), scikit-learn (51), SciPy (42), and Statsmodels (61). The results of this analysis are shown in Datasets S5, S6 and S7.

SI Text Section 4: DDMut Analysis of Platyrrhini-Specific OTR Substitutions.

4.1. Structural Preparation for DDMut Analysis. The structural model used for DDMut analysis was derived from the experimentally determined structure of the human oxytocin receptor (OTR) in complex with oxytocin and a magnesium ion (Mg^{2+}), obtained from the Protein Data Bank (PDB; https://www.rcsb.org) under accession code 7RYC (36). This structure was determined by electron microscopy at a resolution of 3.90 Å, and its validation indicated no deviations from Ramachandran plot parameters (36, 80).

The original 7RYC structure contains unresolved regions in the receptor sequence, specifically residues 1–30, residue 68, residues 237–265, and residues 346–389. To obtain a complete receptor coordinate model for stability prediction, these missing regions were reconstructed using the Robetta server (https://robetta.bakerlab.org) (81, 82). The modeled segments were subsequently evaluated to ensure structural integrity and reliability. The resulting completed OTR–oxytocin–Mg^{2+} coordinate model was used as input for DDMut to estimate the effects of amino acid substitutions on protein stability.

4.2. DDMut Stability Prediction Procedure. The effects of substitutions at positions 169 and 172 of OTR were estimated using DDMut (https://biosig.lab.uq.edu.au/ddmut/) (83), a deep-learning method that predicts changes in protein folding free energy upon mutation ($\Delta\Delta G$). Positive $\Delta\Delta G$ values indicate predicted stabilisation, whereas negative values indicate predicted destabilisation. DDMut uses structural and sequence-derived descriptors from the local residue environment and processes

these features through a siamese neural-network architecture that considers both forward and hypothetical reverse mutations (83).

The completed OTR coordinate model was uploaded to the DDMut web server as a user-provided PDB file. Three mutational scenarios were evaluated: Ala169Val, Val172Met, and the combined Ala169Val/Val172Met double mutant. The two individual substitutions were analysed using the single-mutation option, whereas the combined substitution was analysed using the multiple-mutation option, which is validated for double and triple point mutations (83). The predicted values were used to compare the effect of the double mutant with the additive expectation obtained from the sum of the two single-mutant predictions.

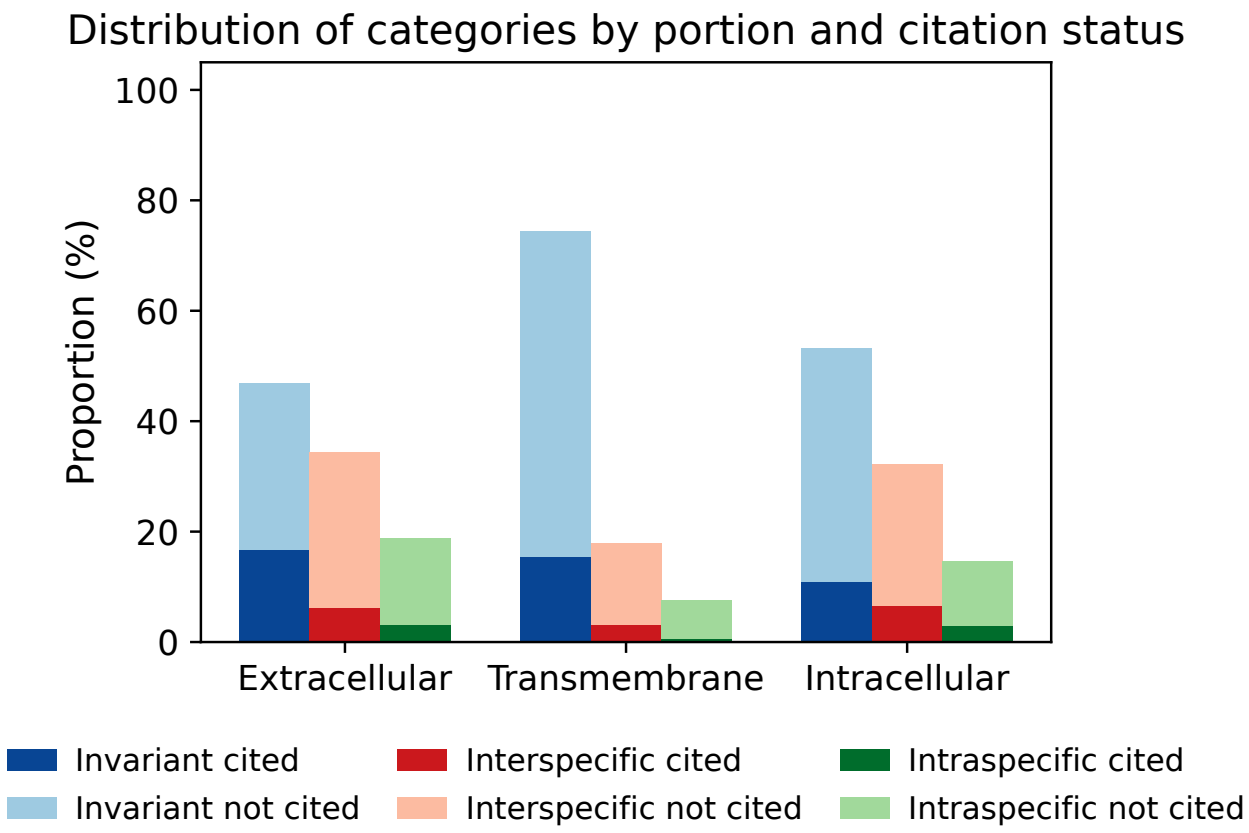
4.3. DDMut Results. DDMut predicted opposite stability effects for the individual and combined substitutions. Ala169Val and Val172Met were individually predicted to be slightly destabilising, with $\Delta\Delta G$ values of -0.01 and -0.30 kcal·mol^{-1}, respectively, resulting in an additive expectation of -0.31 kcal·mol^{-1}. By contrast, the Ala169Val/Val172Met double mutant was predicted to be stabilising, with a $\Delta\Delta G$ of $+1.20$ kcal·mol^{-1} Table 22. This difference corresponds to a positive epistatic effect of approximately $+1.51$ kcal·mol^{-1}, indicating that the combined Val169/Met172 state has a non-additive effect on predicted receptor stability. The same qualitative pattern was also recovered when DDMut was run on the original experimentally resolved OTR structures deposited in the PDB, 7RYC (36) and 7QVM (30), with only minor numerical variation among structural inputs. For this reason, we report the values obtained from the completed Robetta-refined 7RYC-based model, which was used as the main coordinate model for the analysis.

The DDMut output also reported an average distance of 5.51 Å between residues 169 and 172 in TM4 and generated the local interaction map shown in Fig. 3. Together, these results support a model in which the co-occurrence of Val169 and Met172 in Platyrrhini contributes to a TM4 structural background that is more stable than expected from the isolated effects of each substitution. Because these residues are adjacent to Gln171, a conserved residue positioned at the bottom of the ligand-binding pocket, this compensatory stability pattern provides a plausible structural basis for the hypothesis that the Platyrrhini-specific TM4 configuration may have helped preserve receptor stability while allowing diversification of oxytocin ligands.

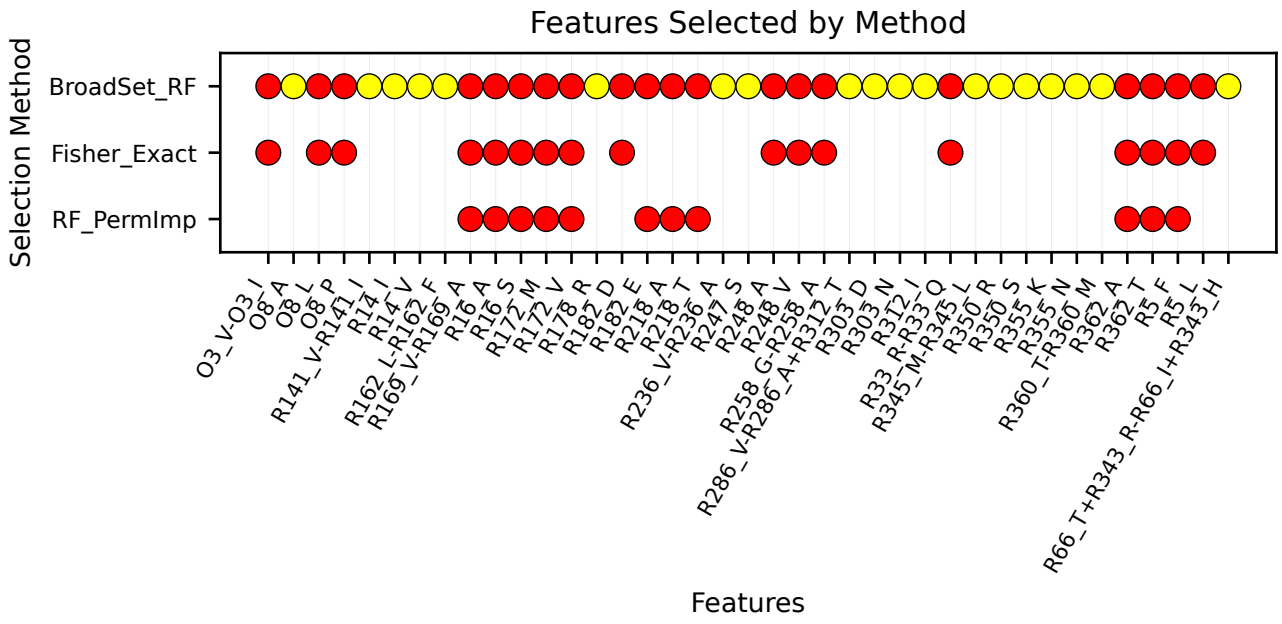
The experiment performed in this work has some limitations regarding the testing of our hypothesis. DDMut estimates effects on protein structural stability and does not directly evaluate ligand binding, receptor activation, downstream signalling, or functional permissivity to ligand variation. Thus, increased ligand permissivity is proposed here as a mechanistic hypothesis supported by compensatory stability predictions, rather than as a direct functional measurement. In addition, because experimentally resolved OTR structures from New World Monkeys are not available, the analyses relied on human OTR structural templates, which may influence estimates of mutational effects in a Platyrrhini sequence background (84). A more direct test of our hypothesis will require experimental approaches beyond the scope of the present study, including functional assays in Platyrrhini OTR backgrounds, comparison among different oxytocin ligands, and receptor activation assays such as BRET-based signalling experiments.

ChatGPT was used to improve the English language of the manuscript and to optimize Python scripts. All outputs were thoroughly reviewed and verified by the authors.

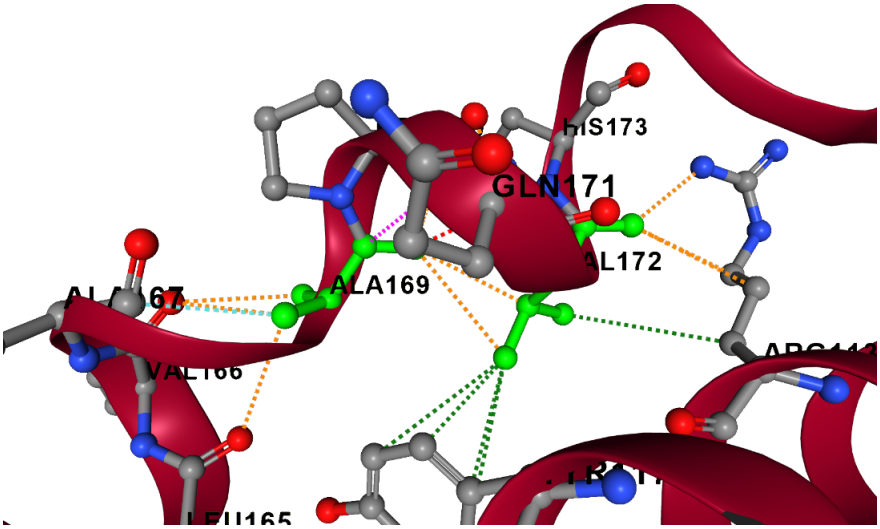
Supplementary Figures



Supplementary Fig. 1. OT and OTR variation. This bar chart shows the distribution of three categories (Invariant, Interspecific, and Intraspecific) in each portion of the protein (Extracellular, Transmembrane, and Intracellular). For each portion, three grouped bars represent the percentage of positions belonging to each category, so that the three bars together total 100%. The blue bars correspond to Invariant positions, the red bars to Interspecific positions, and the green bars to Intraspecific positions. Within each bar, the darker shade indicates the fraction of positions that were cited in the literature, and the lighter shade indicates those that were not cited. The y-axis (Proportion, %) ranges from 0 to 100, enabling a direct comparison of how each portion is composed of the different categories and how often each category has been cited.



Supplementary Fig. 2. Features selected by supervised methods. The figure displays the 39 features selected by at least one method ('R362_A', 'R182_D', 'R182_E', 'R16_A', 'R16_S', 'R362_T', 'O8_A', 'R162_L-R162_F', 'R345_M-R345_L', 'R350_R', 'R172_M', 'R355_N', 'O8_P', 'R350_S', 'R218_A', 'R218_T', 'R172_V', 'R355_K', 'R248_V', 'R360_T-R360_M', 'R258_G-R258_A', 'R5_F', 'O3_V-O3_I', 'R286_V-R286_A+R312_T', 'R169_V-R169_A', 'R14_V', 'R303_N', 'R33_R-R33_Q', 'O8_L', 'R248_A', 'R236_V-R236_A', 'R312_I', 'R5_L', 'R247_S', 'R14_I', 'R66_T+R343_R-R66_I+R343_H', 'R303_D', 'R141_V-R141_I', 'R178_R'). Features shown in yellow were selected exclusively by BroadSet_RF, which defines a comprehensive feature set based on phenotypic classification. Variants in red were also identified by at least one association test (Fisher_Exact and/or RF_PermImp). No variants were identified by PhyloLogReg.



Supplementary Fig. 3. Local interaction environment surrounding residues 169 and 172 in TM4 of the oxytocin receptor. Structural visualization generated in the DDMut interface showing the atomic interaction environment of residues Ala169 and Val172 in the input OTR structure used for stability prediction. The protein backbone is shown in cartoon representation and side chains as sticks. Colored dashed lines indicate interatomic contacts with neighbouring residues identified by DDMut. The visualization represents the local structural context of the reference residues before mutation, whereas the submitted substitutions Ala169Val and Val172Met were used by DDMut to estimate changes in protein stability. This local interaction map provides structural context for the non-additive stability pattern inferred for the combined Ala169Val/Val172Met mutant.

Supplementary Tables

Supplementary Table 1. Reference sequences used for *OXT* and *OXTR* retrieval across primate species.

Species	Ref_seq_*OXT*	Ref_seq_*OXTR*
Allenopithecus nigroviridis	GCA_963574245.1	GCA_963574245.1
Allochrocebus lhoesti	GCA_963574325.1	GCA_963574325.1
Allochrocebus preussi	GCA_963574925.1	GCA_963574925.1
Allochrocebus solatus	GCA_963575195.1	GCA_963575195.1
Alouatta belzebul	GCA_963575095.1 / GCA_963574145.1	GCA_963574145.1 / GCA_963575095.1
Alouatta caraya	GCA_963575295.1 / (85)	KT182966.1 / GCA_963575295.1
Alouatta clamitans	(1)	–
Alouatta discolor	GCA_963575205.1 / (1)	GCA_963575205.1
Alouatta guariba	(1)	–
Alouatta juara	GCA_963574365.1	GCA_963574365.1
Alouatta macconnelli	GCA_963573945.1	GCA_963573945.1
Alouatta palliata	SRX6727374	GCA_963573975.1 / SRX6727375
Alouatta seniculus	GCA_963573765.1	GCA_963573765.1
Alouatta seniculus puruensis	GCA_963574235.1	GCA_963574235.1
Alouatta ululata	(1)	–
Aotus azarae	GCA_963573465.1 / (85)	KT182982.1 / GCA_963573465.1
Aotus griseimembra	GCA_963574505.1	GCA_963574505.1
Aotus lemurinus	–	(85)
Aotus nancymaae	XM_012454997.1	XM_012471062.1
Aotus nigriceps	(1)	–
Aotus trivirgatus	GCA_963574785.1	GCA_963574785.1
Aotus vociferans	GCA_963573775.1	GCA_963573775.1
Arctocebus calabarensis	–	GCA_963573935.1
Ateles belzebuth	GCA_963573605.1 / (85)	KT182968.1 / GCA_963573605.1
Ateles chamek	GCA_963574935.1	GCA_963574935.1
Ateles geoffroyi	GCA_963574205.1 / GCA_004024785.1	GCA_004024785.1 / GCA_963574205.1
Ateles marginatus	GCA_963573555.1	GCA_963573555.1
Ateles paniscus	GCA_963573985.1 / (1)	GCA_963573985.1
Avahi laniger	GCA_963575035.1	GCA_963575035.1
Avahi peyrierasi	GCA_963573525.1	GCA_963573525.1
Brachyteles arachnoides	(1)	–
Brachyteles hypoxanthus	(85)	KT182963.1
Cacajao ayresi	GCA_963573425.1	GCA_963573425.1
Cacajao calvus	GCA_963573955.1 / (85)	KT182972.1 / GCA_963573955.1
Cacajao hosomi	GCA_963573795.1	GCA_963573795.1
Cacajao melanocephalus	GCA_963573385.1 / KM186275	KM186287 / GCA_963573385.1
Callicebus personatus	(1)	–
Callimico goeldii	GCA_963573925.1 / KM186267	GCA_963573925.1 / KM186281
Callithrix aurita	(1)	–
Callithrix geoffroyi	GCA_963573515.1 / KM186262	KM186278 / GCA_963573515.1
Callithrix jacchus	XM_002747304.3	XM_017965047.2
Callithrix kuhlii	GCA_963573495.1 / (85)	GCA_963573495.1 / KT182976.1
Callithrix penicillata	(85)	KT182978.1
Cebuella niveiventris	GCA_963575025.1	GCA_963575025.1
Cebuella pygmaea	(1) / GCA_963575105.1	GCA_963575105.1 / KM186280
Cebus albifrons	SRX8006780	SRX8006780
Cebus capucinus	XM_017504852.2	XM_017501600.2
Cebus imitator	GCF_001604975.1	GCF_001604975.1
Cebus olivaceus	GCA_963574745.1	GCA_963574745.1
Cebus unicolor	–	GCA_963574255.1
Cephalopachus bancanus	GCA_963574495.1	GCA_963574495.1
Cercocebus atys	XM_012052261.1	XM_012059968.1
Cercocebus atys lunulatus	GCA_963573365.1	GCA_963573365.1
Cercocebus chrysogaster	GCA_963574065.1	GCA_963574065.1
Cercocebus torquatus	GCA_963574135.1	GCA_963574135.1
Cercopithecus ascanius	GCA_963574115.1	GCA_963574115.1
Cercopithecus campbelli lowei	GCA_963574615.1	GCA_963574615.1
Cercopithecus cephus	GCA_963575225.1	GCA_963575225.1
Cercopithecus diana	GCA_963574765.1	GCA_963574765.1
Cercopithecus hamlyni	GCA_963574815.1	GCA_963574815.1
Cercopithecus neglectus	GCA_963574685.1 / SRX8006877	GCA_963574685.1 / SRX8006877
Cercopithecus nictitans	GCA_963575045.1	GCA_963575045.1
Cercopithecus petaurista	GCA_963575065.1	GCA_963575065.1
Cercopithecus pogonias	GCA_963575055.1	GCA_963575055.1
Cercopithecus roloway	GCA_963573595.1	GCA_963573595.1
Cheirogaleus major	GCA_963574435.1	GCA_963574435.1
Cheirogaleus medius	SRX8655184 / GCA_963574425.1	GCA_963574425.1 / SRX8655184
Cheracebus lucifer	GCA_963573755.1	GCA_963573755.1

Continued on next page

Table S1. Continued.

Species	Ref_seq_*OXT*	Ref_seq_*OXTR*
Cheracebus lugens	GCA_963574535.1	GCA_963574535.1
Cheracebus regulus	GCA_963574635.1	GCA_963574635.1
Cheracebus torquatus	GCA_963574645.1	GCA_963574645.1
Chiropotes albinasus	KM186277	KM186289
Chiropotes chiropotes	(85)	KT182971.1
Chiropotes sagulatus	GCA_963574175.1	GCA_963574175.1
Chiropotes satanas	(1)	–
Chiropotes utahickae	KM186276	KM186288
Chlorocebus pygerythrus	GCA_963575115.1	GCA_963575115.1
Chlorocebus sabaeus	XM_008019220.2	XM_007985098.2
Colobus angolensis	–	XM_011961277.1
Colobus polykomos	GCA_963574725.1	GCA_963574725.1
Daubentonia madagascariensis	GCA_004027145.1	GCA_004027145.1
Erythrocebus patas	SRS6382159	SRS6382159
Eulemur albifrons	GCA_963573435.1	GCA_963573435.1
Eulemur coronatus	GCA_963574965.1	GCA_963574965.1
Eulemur flavifrons	SRS945120 / GCA_963574995.1	SRS945120 / GCA_963574995.1
Eulemur fulvus	GCA_004027275.1 / GCA_963573725.1	GCA_004027275.1 / GCA_963573725.1
Eulemur fulvus collaris	GCA_963575015.1	GCA_963575015.1
Eulemur macaco	GCA_963574945.1	GCA_963574945.1
Eulemur mongoz	GCA_963574475.1	GCA_963574475.1
Eulemur rubriventer	GCA_963574895.1	GCA_963574895.1
Eulemur rufus	GCA_963574465.1	GCA_963574465.1
Eulemur sanfordi	GCA_963573705.1	GCA_963573705.1
Galago senegalensis	GCA_963575145.1	GCA_963575145.1
Galagoides demidoff	GCA_963574575.1	GCA_963574575.1
Gorilla beringei	GCA_963575185.1	GCA_963575185.1
Gorilla gorilla	XM_031004971.1	XM_031009515.1
Hapalemur griseus	GCA_963575165.1 / GCA_963573835.1	GCA_963573835.1 / GCA_963575165.1
Hapalemur griseus alaotrensis	GCA_963574485.1	GCA_963574485.1
Hapalemur meridionalis	GCA_963575175.1	GCA_963575175.1
Hapalemur occidentalis	GCA_963575155.1	GCA_963575155.1
Homo sapiens	NM_000915.4	NM_000916.4
Hylobates agilis	GCA_963574095.1	GCA_963574095.1
Hylobates klossii	GCA_963574335.1	GCA_963574335.1
Hylobates moloch	XM_032142588.1	XM_032149565.1
Hylobates muelleri	GCA_963573745.1	GCA_963573745.1
Hylobates muelleri abbotti	GCA_963574865.1	GCA_963574865.1
Indri indri	GCA_963575135.1 / SRX8008840	GCA_963575135.1 / SRX8008840
Lagothrix lagotricha	GCA_963574225.1 / (85)	GCA_963574225.1 / KT182964.1
Lagothrix poeppigii	(85)	KT182965.1
Lemur catta	GCA_004024665.1	GCA_004024665.1
Leontocebus fuscicollis	GCA_963575285.1 / GCA_963573655.1	GCA_963575285.1 / GCA_963573655.1
Leontocebus nigricollis	GCA_963575265.1	GCA_963575265.1
Leontopithecus chrysomelas	GCA_963574395.1 / KM186268	GCA_963574395.1
Leontopithecus chrysopygas	(1)	–
Leontopithecus rosalia	GCA_963574445.1 / KM186269	GCA_963574445.1 / KT182980.1
Lepilemur ankaranensis	GCA_963573785.1	GCA_963573785.1
Lepilemur dorsalis	GCA_963574875.1	GCA_963574875.1
Lepilemur ruficaudatus	GCA_963574885.1	GCA_963574885.1
Lepilemur septentrionalis	GCA_963573855.1	GCA_963573855.1
Loris lydekkerianus	GCA_963574355.1	GCA_963574355.1
Macaca arctoides	GCA_963573575.1	GCA_963573575.1
Macaca cyclopis	GCA_963573865.1	GCA_963573865.1
Macaca fascicularis	GCA_963574375.1 / NM_001283971.1	GCA_963574375.1 / NM_001319545.1
Macaca fuscata	GCA_003118495.1	GCA_003118495.1
Macaca fuscata fuscata	GCA_963573635.1	GCA_963573635.1
Macaca leonina	GCA_963573825.1	GCA_963573825.1
Macaca maura	GCA_963574855.1	GCA_963574855.1
Macaca mulatta	XM_001115045.4	NM_001044732.1
Macaca nemestrina	XM_011741312.2	XM_011734209.1
Macaca nigra	GCA_963573485.1	GCA_963573485.1
Macaca radiata	GCA_963573735.1	GCA_963573735.1
Macaca siberu	GCA_963574275.1	GCA_963574275.1
Macaca thibetana	GCA_963574025.1	GCA_963574025.1
Macaca tonkeana	GCA_963573625.1	GCA_963573625.1
Mandrillus leucophaeus	XM_011974736.1	XM_012000637.1
Mandrillus sphinx	SRS3024194	SRS3024194
Mico argentatus	(85) / GCA_963573845.1 / GCA_963573565.1	GCA_963573845.1 / GCA_963573565.1
Mico chrysoleucus	(1)	–
Mico humeralifer	GCA_963574605.1 / KM186265	GCA_963574605.1 / KM186279

Continued on next page

Table S1. Continued.

Species	Ref_seq_*OXT*	Ref_seq_*OXTR*
Mico humilis	(1) / GCA_963574545.1	GCA_963574545.1
Mico mauesi	(1)	–
Mico saterei	KM186264	–
Microcebus griseorufus	SRX5263178	SRX5263178
Microcebus mittermeieri	SRX5263177	SRX5263177
Microcebus murinus	XM_012754916.1	XM_020284221.1
Microcebus ravelobensis	SRX5263180	SRX5263180
Microcebus tavaratra	SRX5263182	SRX5263182
Miopithecus ogouensis	GCA_963573445.1	GCA_963573445.1
Mirza coquereli	GCA_004024645.1	GCA_004024645.1
Mirza zaza	GCA_963573685.1 / SRX5263181	GCA_963573685.1 / SRX5263181
Nasalis larvatus	GCA_963574835.1 / SRS6383564	GCA_963574835.1 / SRS6383564
Nomascus annamensis	GCA_963574525.1	GCA_963574525.1
Nomascus concolor	GCA_963574845.1	GCA_963574845.1
Nomascus gabriellae	GCA_963574555.1	GCA_963574555.1
Nomascus leucogenys	XM_003277949.3	XM_030801813.1
Nomascus siki	GCA_963573715.1	GCA_963573715.1
Nycticebus coucang	GCA_963573995.1 / SRX8010127	GCA_963573995.1 / SRX8010127
Otolemur crassicaudatus	GCA_963574585.1	GCA_963574585.1
Otolemur garnettii	XM_003788219.1	XM_003785470.1
Pan paniscus	XM_034947010.1	XM_034955695.1
Pan troglodytes	XM_001160221.6	XM_001144020.6
Papio anubis	XM_017945525.3	XM_021934220.2
Papio cynocephalus	GCA_963574715.1	GCA_963574715.1
Papio kindae	GCA_963573405.1	GCA_963573405.1
Papio papio	GCA_963573375.1	GCA_963573375.1
Papio ursinus	GCA_963574515.1	GCA_963574515.1
Perodicticus potto	GCA_963574655.1	GCA_963574655.1
Perodicticus potto ibeanus	GCA_963574675.1	GCA_963574675.1
Piliocolobus badius	GCA_963574265.1	GCA_963574265.1
Piliocolobus gordonorum	GCA_963574285.1	GCA_963574285.1
Piliocolobus kirkii	GCA_963574805.1	GCA_963574805.1
Piliocolobus tephrosceles	XM_031934673.1	XM_023218386.2
Pithecia albicans	GCA_963573695.1	GCA_963573695.1
Pithecia chrysocephala	GCA_963574905.1	GCA_963574905.1
Pithecia hirsuta	GCA_963573645.1	GCA_963573645.1
Pithecia mittermeier	(1)	–
Pithecia mittermeieri	GCA_963574975.1	GCA_963574975.1
Pithecia pissinatti	GCA_963575305.1	GCA_963575305.1
Pithecia pithecia	(85)	SRX8010141
Pithecia vanzolinii	GCA_963574005.1	GCA_963574005.1
Plecturocebus bernhardi	GCA_963574705.1	GCA_963574705.1
Plecturocebus brunneus	GCA_963573885.1	GCA_963573885.1
Plecturocebus caligatus	GCA_963575255.1 / (1)	GCA_963575255.1
Plecturocebus caquetensis	(1)	–
Plecturocebus cinerascens	GCA_963573415.1	GCA_963573415.1
Plecturocebus cupreus	GCA_963574565.1 / JF315862.1	KT182969.1 / GCA_963574565.1
Plecturocebus donacophilus	(1) / SRX8006773	SRX8006773
Plecturocebus dubius	GCA_963575235.1 / (1)	GCA_963575235.1
Plecturocebus grovesi	GCA_963574625.1	GCA_963574625.1
Plecturocebus hoffmannsi	GCA_963575245.1	GCA_963575245.1
Plecturocebus miltoni	GCA_963575275.1	GCA_963575275.1
Plecturocebus moloch	GCA_963574105.1	GCA_963574105.1 / (85)
Pongo abelii	XM_002830099.2	XM_002813482.3
Pongo pygmaeus	GCF_028885625.2	GCA_900086635.1
Presbytis comata	GCA_963573905.1	GCA_963573905.1
Presbytis melalophos mitrata	GCA_963575215.1	GCA_963575215.1
Prolemur simus	SRX2229843	SRX2229843
Propithecus coquereli	XM_012665739.1	XM_012646708.1
Propithecus deckenii coronatus	GCA_963574665.1	GCA_963574665.1
Propithecus diadema	GCA_963575075.1	GCA_963575075.1
Propithecus edwardsi	GCA_963575085.1	GCA_963575085.1
Propithecus perrieri	GCA_963574695.1	GCA_963574695.1
Propithecus tattersalli	GCA_963574315.1	GCA_963574315.1
Propithecus verreauxi	GCA_963575125.1	GCA_963575125.1
Pygathrix cinerea	GCA_963574085.1	GCA_963574085.1
Pygathrix nemaeus	GCA_004024825.1	GCA_004024825.1
Pygathrix nigripes	GCA_963574055.1	GCA_963574055.1
Rhinopithecus bieti	GCA_963573815.1 / XM_017872529.1	GCA_963573815.1 / XM_017874159.1
Rhinopithecus roxellana	XM_010378981.2	XM_010369904.2

Continued on next page

Table S1. Continued.

Species	Ref_seq_*OXT*	Ref_seq_*OXTR*
Saguinus bicolor	KM186272 / GCA_963574455.1 / GCA_963574165.1	KM186282 / GCA_963574455.1 / GCA_963574165.1
Saguinus geoffroyi	GCA_963574215.1	GCA_963574215.1
Saguinus imperator	GCA_963574955.1 / GCA_004024885.1	GCA_963574955.1 / GCA_004024885.1
Saguinus inustus	GCA_963574735.1	GCA_963574735.1
Saguinus labiatus	GCA_963575005.1	GCA_963575005.1
Saguinus martinsi	KM186271	KM186283
Saguinus midas	(1)/SRR096434	KT182981.1
Saguinus mystax	GCA_963574345.1	GCA_963574345.1
Saguinus niger	KM186270	KM186284
Saguinus oedipus	GCA_963574915.1	GCA_963574915.1
Saimiri boliviensis	XM_003941063.1	XM_003927113.3
Saimiri cassiquiarensis	GCA_963573395.1	GCA_963573395.1
Saimiri macrodon	GCA_963573505.1	GCA_963573505.1
Saimiri oerstedii	GCA_963574195.1	GCA_963574195.1
Saimiri sciureus	GCA_963573345.1 / JF_315866.1	JF_330026.1 / GCA_963573345.1
Saimiri ustus	GCA_963574015.1	GCA_963574015.1
Sapajus apella	XM_032294909.1	XM_032290544.1
Sapajus apella macrocephalus	GCA_963574755.1	GCA_963574755.1
Sapajus libidinosus	(1)	–
Sapajus robustus	KM186273	KM186285
Sapajus xanthosternos	KM186274	KM186286
Semnopithecus entellus	GCA_963573915.1 / SRX8012806	SRX8012806 / GCA_963573915.1
Semnopithecus hypoleucos	GCA_963573665.1	GCA_963573665.1
Semnopithecus priam	GCA_963574295.1	GCA_963574295.1
Semnopithecus schistaceus	GCA_963574385.1	GCA_963574385.1
Symphalangus syndactylus	XM_055264935.1	XM_055259247.1
Tarsius lariang	GCA_963574405.1	GCA_963574405.1
Tarsius wallacei	GCA_963574415.1	GCA_963574415.1
Theropithecus gelada	XM_025400051.1	XM_025377661.1
Trachypithecus auratus	GCA_963574155.1	GCA_963574155.1
Trachypithecus cristatus	GCA_963574045.1	GCA_963574045.1
Trachypithecus francoisi	GCA_963574775.1 / XM_033185855.1	GCA_963574775.1 / XM_033210524.1
Trachypithecus geei	GCA_963573545.1	GCA_963573545.1
Trachypithecus germaini	GCA_963573455.1	GCA_963573455.1
Trachypithecus hatinhensis	GCA_963573355.1	GCA_963573355.1
Trachypithecus johnii	GCA_963574185.1	GCA_963574185.1
Trachypithecus laotum	GCA_963573585.1	GCA_963573585.1
Trachypithecus obscurus	GCA_963574075.1	GCA_963574075.1
Trachypithecus phayrei	GCA_963573535.1	GCA_963573535.1
Trachypithecus phayrei crepuscula	GCA_963574125.1	GCA_963574125.1
Trachypithecus pileatus	GCA_963573805.1	GCA_963573805.1
Trachypithecus poliocephalus	GCA_963573475.1	GCA_963573475.1
Trachypithecus vetulus	GCA_963574985.1	GCA_963574985.1
Varecia rubra	GCA_963573675.1	GCA_963573675.1

Supplementary Table 2. Social monogamy phenotype classification across primate species and the corresponding bibliographic sources.

Family	Species	Phenotype_SM	Ref. social monogamy
Lorisidae	*Perodicticus potto ibeanus*	0	(25); (4)
Lorisidae	*Nycticebus coucang*	0	(86); (24); (20); (25); (4)
Lorisidae	*Loris lydekkerianus*	0	(87); (20); (4)
Galagidae	*Galagoides demidoff*	0	(4)
Galagidae	*Galago senegalensis*	0	(20); (4)
Galagidae	*Otolemur garnettii*	0	(24); (20); (25)
Galagidae	*Otolemur crassicaudatus*	0	(20); (4)
Daubentoniidae	*Daubentonia madagascariensis*	0	(24); (20); (25)
Indriidae	*Avahi peyrierasi*	1	(25)
Indriidae	*Avahi laniger*	1	(20); (4)
Indriidae	*Propithecus deckenii coronatus*	0	(88); (20)
Indriidae	*Propithecus verreauxi*	0	(20); (4)
Indriidae	*Propithecus coquereli*	0	(24)*; (20); (25)*; (4)
Indriidae	*Propithecus tattersalli*	0	(20); (4)
Indriidae	*Propithecus edwardsi*	0	(20); (4)
Indriidae	*Propithecus diadema*	0	(89); (20); (4)
Indriidae	*Propithecus perrieri*	0	(88); (20)
Indriidae	*Indri indri*	1	(24); (20); (27); (25); (26)

Continued on next page

Table S2. Continued.

Family	Species	Phenotype_SM	Ref. social monogamy
Cheirogaleidae	*Mirza coquereli*	0	(24); (20); (25)
Cheirogaleidae	*Mirza zaza*	0	(90); (25)
Cheirogaleidae	*Microcebus mittermeieri*	0	(91); (25)
Cheirogaleidae	*Microcebus tavaratra*	0	(91); (25)
Cheirogaleidae	*Microcebus ravelobensis*	0	(24); (20); (91); (25)
Cheirogaleidae	*Microcebus griseorufus*	0	(20); (25)
Cheirogaleidae	*Microcebus murinus*	0	(20); (4)
Cheirogaleidae	*Cheirogaleus major*	1	(92); (20)
Cheirogaleidae	*Cheirogaleus medius*	1	(24); (20); (27)*; (91); (26)
Lepilemuridae	*Lepilemur ankaranensis*	0	(20); (4)
Lepilemuridae	*Lepilemur dorsalis*	0	(20); (4)
Lepilemuridae	*Lepilemur septentrionalis*	0	(20); (4)
Lepilemuridae	*Lepilemur ruficaudatus*	0	(20); (4)
Lemuridae	*Varecia rubra*	0	(20); (4)
Lemuridae	*Prolemur simus*	0	(24); (20); (25)
Lemuridae	*Hapalemur occidentalis*	1	(24); (25); (4)
Lemuridae	*Lemur catta*	0	(24); (20); (25)
Lemuridae	*Eulemur mongoz*	1	(24)*; (20); (25)*
Lemuridae	*Eulemur rubriventer*	1	(24); (20); (25)
Lemuridae	*Eulemur rufus*	0	(20); (25)
Lemuridae	*Eulemur albifrons*	0	(24); (25); (4)
Lemuridae	*Eulemur sanfordi*	0	(24); (25); (4)
Lemuridae	*Eulemur flavifrons*	0	(24); (25); (4)
Lemuridae	*Eulemur macaco*	0	(24); (20); (25)*; (4)
Lemuridae	*Eulemur coronatus*	0	(24); (20); (25); (4)
Pitheciidae	*Pithecia albicans*	1	(20)
Pitheciidae	*Pithecia pithecia*	1	(20); (25)*; (26)
Pitheciidae	*Cacajao calvus*	0	(20); (4)
Pitheciidae	*Cacajao hosomi*	0	(25)
Pitheciidae	*Cacajao ayresi*	0	(25)
Pitheciidae	*Cacajao melanocephalus*	0	(20); (25)
Pitheciidae	*Chiropotes albinasus*	0	(20)
Pitheciidae	*Chiropotes chiropotes*	0	(20)
Pitheciidae	*Chiropotes utahickae*	0	(20); (25)
Pitheciidae	*Cheracebus regulus*	1	(20)
Pitheciidae	*Cheracebus torquatus*	1	(20); (4)
Pitheciidae	*Cheracebus lucifer*	1	(20)
Pitheciidae	*Cheracebus lugens*	1	(20)
Pitheciidae	*Plecturocebus cinerascens*	1	(20)
Pitheciidae	*Plecturocebus moloch*	1	(24); (20); (25); (4)
Pitheciidae	*Plecturocebus bernhardi*	1	(20)
Pitheciidae	*Plecturocebus hoffmannsi*	1	(20)
Pitheciidae	*Plecturocebus brunneus*	1	(20)
Pitheciidae	*Plecturocebus cupreus*	1	(20); (25)
Pitheciidae	*Plecturocebus dubius*	1	(20)
Pitheciidae	*Plecturocebus caligatus*	1	(20); (25)
Pitheciidae	*Plecturocebus donacophilus*	1	(24); (20); (27); (25)
Aotidae	*Aotus nigriceps*	1	(20); (25); (26)
Aotidae	*Aotus azarai*	1	(24); (20); (25); (4)
Aotidae	*Aotus griseimembra*	1	(25)
Aotidae	*Aotus trivirgatus*	1	(24); (20); (25); (4)
Aotidae	*Aotus vociferans*	1	(20); (25)
Aotidae	*Aotus nancymaae*	1	(24); (20); (25)
Callitrichidae	*Mico humeralifer*	1	(24)*; (20); (27)*
Callitrichidae	*Mico argentatus*	1	(4)
Callitrichidae	*Cebuella pygmaea*	1	(24)*; (20); (27)
Callitrichidae	*Mico humilis*	1	(20)
Callitrichidae	*Callithrix penicillata*	1	(24)*; (20)
Callitrichidae	*Callithrix jacchus*	1	(24)*; (20); (27)*
Callitrichidae	*Callithrix kuhlii*	1	(24)*; (20)
Callitrichidae	*Callithrix geoffroyi*	1	(24)*; (20); (27)*
Callitrichidae	*Callimico goeldii*	1	(93)*; (94); (24)*; (95); (27)*; (47); (96); (21)
Callitrichidae	*Leontopithecus chrysomelas*	1	(24)*; (20); (25)*; (4)
Callitrichidae	*Leontopithecus rosalia*	1	(24)*; (20); (27)*; (26)
Callitrichidae	*Leontocebus nigricollis*	1	(20)
Callitrichidae	*Saguinus mystax*	1	(20)
Callitrichidae	*Saguinus labiatus*	1	(20)
Callitrichidae	*Saguinus imperator*	1	(24)*; (20); (27)*
Callitrichidae	*Saguinus midas*	1	(24)*; (20); (27)
Callitrichidae	*Saguinus niger*	1	(20)
Callitrichidae	*Saguinus bicolor*	1	(20)

Continued on next page

Table S2. Continued.

Family	Species	Phenotype_SM	Ref. social monogamy
Callitrichidae	*Saguinus martinsi*	1	(20)
Callitrichidae	*Saguinus oedipus*	1	(20); (4)
Callitrichidae	*Saguinus geoffroyi*	1	(20)
Callitrichidae	*Saguinus inustus*	1	(20)
Cebidae	*Cebus olivaceus*	0	(20); (4)
Cebidae	*Cebus capucinus*	0	(24); (20); (25)
Cebidae	*Cebus albifrons*	0	(24); (20); (25)
Cebidae	*Sapajus xanthosternos*	0	(20); (25)
Cebidae	*Sapajus apella macrocephalus*	0	(20); (4)
Cebidae	*Sapajus robustus*	0	(20); (25); (91)
Cebidae	*Saimiri ustus*	0	(20)
Cebidae	*Saimiri oerstedii*	0	(20)
Cebidae	*Saimiri macrodon*	0	(4)
Cebidae	*Saimiri boliviensis*	0	(24); (20); (25)
Atelidae	*Ateles marginatus*	0	(20)
Atelidae	*Ateles chamek*	0	(20)
Atelidae	*Ateles paniscus*	0	(20); (4)
Atelidae	*Ateles geoffroyi*	0	(24); (20); (25)
Atelidae	*Ateles belzebuth*	0	(20); (4)
Atelidae	*Brachyteles hypoxanthus*	0	(20)
Atelidae	*Lagothrix poeppigii*	0	(20)
Atelidae	*Lagothrix lagotricha*	0	(20)
Atelidae	*Alouatta macconnelli*	0	(20)
Atelidae	*Alouatta seniculus puruensis*	0	(20); (4)
Atelidae	*Alouatta caraya*	0	(20); (4)
Atelidae	*Alouatta belzebul*	0	(20)
Atelidae	*Alouatta palliata*	0	(24); (20); (25)*
Hylobatidae	*Symphalangus syndactylus*	1	(20)
Hylobatidae	*Hylobates agilis*	1	(20); (4)
Hylobatidae	*Hylobates muelleri*	1	(20); (4)
Hylobatidae	*Hylobates moloch*	1	(24); (20); (25); (26)
Hylobatidae	*Hylobates klossii*	1	(20); (4)
Hylobatidae	*Nomascus leucogenys*	1	(97); (24); (27); (25)
Hylobatidae	*Nomascus siki*	1	(25)
Hylobatidae	*Nomascus gabriellae*	1	(4)
Hylobatidae	*Nomascus concolor*	1	(4)
Hominidae	*Gorilla gorilla*	0	(24); (20); (25)
Hominidae	*Gorilla beringei*	0	(20); (4)
Hominidae	*Homo sapiens*	1	(20); (98); (21)
Hominidae	*Pan paniscus*	0	(24); (20); (25)
Hominidae	*Pan troglodytes*	0	(24); (20); (25)
Hominidae	*Pongo abelii*	0	(24); (20); (25)
Cercopithecidae	*Chlorocebus pygerythrus*	0	(4)
Cercopithecidae	*Chlorocebus sabaeus*	0	(24); (20); (25)
Cercopithecidae	*Cercopithecus campbelli lowei*	0	(20)
Cercopithecidae	*Erythrocebus patas*	0	(24); (20); (25)
Cercopithecidae	*Allochrocebus solatus*	0	(4)
Cercopithecidae	*Allochrocebus preussi*	0	(4)
Cercopithecidae	*Allochrocebus lhoesti*	0	(4)
Cercopithecidae	*Cercopithecus ascanius*	0	(20); (4)
Cercopithecidae	*Cercopithecus petaurista*	0	(20); (4)
Cercopithecidae	*Cercopithecus nictitans*	0	(20); (4)
Cercopithecidae	*Cercopithecus diana*	0	(20); (4)
Cercopithecidae	*Cercopithecus roloway*	0	(20)
Cercopithecidae	*Cercopithecus neglectus*	1	(24); (27)*; (25)*; (26)
Cercopithecidae	*Cercopithecus hamlyni*	0	(20); (4)
Cercopithecidae	*Miopithecus ogouensis*	0	(20)
Cercopithecidae	*Allenopithecus nigroviridis*	0	(20); (4)
Cercopithecidae	*Papio papio*	0	(20); (4)
Cercopithecidae	*Papio ursinus*	0	(20); (4)
Cercopithecidae	*Papio cynocephalus*	0	(20); (4)
Cercopithecidae	*Papio anubis*	0	(24); (20); (25)
Cercopithecidae	*Theropithecus gelada*	0	(20); (25)
Cercopithecidae	*Mandrillus sphinx*	0	(24); (20); (25)
Cercopithecidae	*Mandrillus leucophaeus*	0	(24); (20); (25)
Cercopithecidae	*Cercocebus chrysogaster*	0	(20)
Cercopithecidae	*Cercocebus atys lunulatus*	0	(20)
Cercopithecidae	*Cercocebus torquatus*	0	(20); (4)
Cercopithecidae	*Macaca thibetana*	0	(20); (4)
Cercopithecidae	*Macaca radiata*	0	(20); (4)
Cercopithecidae	*Macaca arctoides*	0	(20); (4)

Continued on next page

Table S2. Continued.

Family	Species	Phenotype_SM	Ref. social monogamy
Cercopithecidae	*Macaca mulatta*	0	(24); (20); (25)
Cercopithecidae	*Macaca cyclopis*	0	(20); (4)
Cercopithecidae	*Macaca fuscata fuscata*	0	(20); (4)
Cercopithecidae	*Macaca fascicularis*	0	(24); (20); (25)
Cercopithecidae	*Macaca nemestrina*	0	(24); (20); (25)
Cercopithecidae	*Macaca siberu*	0	(20)
Cercopithecidae	*Macaca leonina*	0	(20); (4)
Cercopithecidae	*Macaca nigra*	0	(20); (4)
Cercopithecidae	*Macaca tonkeana*	0	(20); (4)
Cercopithecidae	*Macaca maura*	0	(20); (4)
Cercopithecidae	*Semnopithecus entellus*	0	(24); (20); (25)
Cercopithecidae	*Trachypithecus johnii*	0	(20)
Cercopithecidae	*Semnopithecus priam*	0	(20)
Cercopithecidae	*Trachypithecus vetulus*	0	(20)
Cercopithecidae	*Trachypithecus obscurus*	0	(20); (4)
Cercopithecidae	*Trachypithecus auratus*	0	(20); (4)
Cercopithecidae	*Trachypithecus cristatus*	0	(20); (4)
Cercopithecidae	*Trachypithecus germaini*	0	(20)
Cercopithecidae	*Trachypithecus francoisi*	0	(24); (20); (25)
Cercopithecidae	*Trachypithecus poliocephalus*	0	(20)
Cercopithecidae	*Trachypithecus hatinhensis*	0	(20)
Cercopithecidae	*Trachypithecus laotum*	0	(20)
Cercopithecidae	*Trachypithecus geei*	0	(20); (4)
Cercopithecidae	*Trachypithecus pileatus*	0	(20); (4)
Cercopithecidae	*Presbytis comata*	0	(20); (4)
Cercopithecidae	*Presbytis melalophos mitrata*	0	(20)
Cercopithecidae	*Nasalis larvatus*	0	(24); (20); (25)*
Cercopithecidae	*Pygathrix nigripes*	0	(20)
Cercopithecidae	*Pygathrix cinerea*	0	(20)
Cercopithecidae	*Pygathrix nemaeus*	0	(24); (20); (25)
Cercopithecidae	*Rhinopithecus roxellana*	0	(24); (20); (25)
Cercopithecidae	*Rhinopithecus bieti*	0	(24); (20); (25)
Cercopithecidae	*Piliocolobus badius*	0	(20); (4)
Cercopithecidae	*Piliocolobus tephrosceles*	0	(20)
Cercopithecidae	*Piliocolobus kirkii*	0	(20)
Cercopithecidae	*Piliocolobus gordonorum*	0	(20)
Cercopithecidae	*Colobus polykomos*	0	(20); (4)
Tarsiidae	*Tarsius lariang*	1	(25)*; (26)
Tarsiidae	*Tarsius wallacei*	1	(25)*
Tarsiidae	*Cephalopachus bancanus*	0	(4)

* Indicates more than one phenotypic classification for the species in the cited reference.

Supplementary Table 3. Distribution of OT peptide variants across primate species, including the number of species, number of samples, and the corresponding species list.

Variant	Species_Count	Samples_Count	Species_List
$Leu^{8}OT$	171	188	*Allenopithecus nigroviridis*, *Allochrocebus lhoesti*, *Allochrocebus preussi*, *Allochrocebus solatus*, *Alouatta clamitans*, *Alouatta discolor*, *Alouatta guariba*, *Alouatta palliata*, *Alouatta ululata*, *Avahi laniger*, *Avahi peyrierasi*, *Brachyteles arachnoides*, *Brachyteles hypoxanthus*, *Callicebus personatus*, *Cephalopachus bancanus*, *Cercocebus atys*, *Cercocebus atys lunulatus*, *Cercocebus chrysogaster*, *Cercocebus torquatus*, *Cercopithecus ascanius*, *Cercopithecus campbelli lowei*, *Cercopithecus cephus*, *Cercopithecus diana*, *Cercopithecus hamlyni*, *Cercopithecus neglectus*, *Cercopithecus nictitans*, *Cercopithecus petaurista*, *Cercopithecus pogonias*, *Cercopithecus roloway*, *Cheirogaleus major*, *Cheirogaleus medius*, *Cheracebus lucifer*, *Cheracebus lugens*, *Cheracebus regulus*, *Cheracebus torquatus*, *Chlorocebus pygerythrus*, *Chlorocebus sabaeus*, *Colobus polykomos*, *Daubentonia madagascariensis*, *Erythrocebus patas*, *Eulemur albifrons*, *Eulemur coronatus*, *Eulemur flavifrons*, *Eulemur fulvus*, *Eulemur fulvus collaris*, *Eulemur macaco*, *Eulemur mongoz*, *Eulemur rubriventer*, *Eulemur rufus*, *Eulemur sanfordi*, *Galago senegalensis*, *Galagoides demidoff*, *Gorilla beringei*, *Gorilla gorilla*, *Hapalemur griseus*, *Hapalemur griseus alaotrensis*, *Hapalemur meridionalis*, *Hapalemur occidentalis*, *Homo sapiens*, *Hylobates agilis*, *Hylobates klossii*, *Hylobates moloch*, *Hylobates muelleri*, *Hylobates muelleri abbotti*, *Indri indri*, *Lagothrix lagotricha*, *Lagothrix poeppigii*, *Lemur catta*, *Lepilemur ankaranensis*, *Lepilemur dorsalis*, *Lepilemur ruficaudatus*, *Lepilemur septentrionalis*, *Loris lydekkerianus*, *Macaca arctoides*, *Macaca cyclopis*, *Macaca fascicularis*, *Macaca fuscata*, *Macaca fuscata fuscata*, *Macaca leonina*, *Macaca maura*, *Macaca mulatta*, *Macaca nemestrina*, *Macaca nigra*, *Macaca radiata*, *Macaca siberu*, *Macaca thibetana*, *Macaca tonkeana*, *Mandrillus leucophaeus*, *Mandrillus sphinx*, *Microcebus griseorufus*, *Microcebus mittermeieri*, *Microcebus murinus*, *Microcebus ravelobensis*, *Microcebus tavaratra*, *Miopithecus ogouensis*, *Mirza coquereli*, *Mirza zaza*, *Nasalis larvatus*, *Nomascus annamensis*, *Nomascus concolor*, *Nomascus gabriellae*, *Nomascus leucogenys*, *Nomascus siki*, *Nycticebus coucang*, *Otolemur crassicaudatus*, *Otolemur garnettii*, *Pan paniscus*, *Pan troglodytes*, *Papio anubis*, *Papio cynocephalus*, *Papio kindae*, *Papio papio*, *Papio ursinus*, *Perodicticus potto*, *Perodicticus potto ibeanus*, *Piliocolobus badius*, *Piliocolobus gordonorum*, *Piliocolobus kirkii*, *Piliocolobus tephrosceles*, *Plecturocebus bernhardi*, *Plecturocebus brunneus*, *Plecturocebus caligatus*, *Plecturocebus caquetensis*, *Plecturocebus cinerascens*, *Plecturocebus cupreus*, *Plecturocebus donacophilus*, *Plecturocebus dubius*, *Plecturocebus grovesi*, *Plecturocebus hoffmannsi*, *Plecturocebus miltoni*, *Plecturocebus moloch*, *Pongo abelii*, *Pongo pygmaeus*, *Presbytis comata*, *Presbytis melalophos mitrata*, *Prolemur simus*, *Propithecus coquereli*, *Propithecus deckenii coronatus*, *Propithecus diadema*, *Propithecus edwardsi*, *Propithecus perrieri*, *Propithecus tattersalli*, *Propithecus verreauxi*, *Pygathrix cinerea*, *Pygathrix nemaeus*, *Pygathrix nigripes*, *Rhinopithecus bieti*, *Rhinopithecus roxellana*, *Semnopithecus entellus*, *Semnopithecus hypoleucos*, *Semnopithecus priam*, *Semnopithecus schistaceus*, *Symphalangus syndactylus*, *Tarsius lariang*, *Tarsius wallacei*, *Theropithecus gelada*, *Trachypithecus auratus*, *Trachypithecus cristatus*, *Trachypithecus francoisi*, *Trachypithecus geei*, *Trachypithecus germaini*, *Trachypithecus hatinhensis*, *Trachypithecus johnii*, *Trachypithecus laotum*, *Trachypithecus obscurus*, *Trachypithecus phayrei*, *Trachypithecus phayrei crepuscula*, *Trachypithecus pileatus*, *Trachypithecus poliocephalus*, *Trachypithecus vetulus*, *Varecia rubra*
$Phe^{2}OT$	4	6	*Alouatta belzebul*, *Alouatta caraya*, *Alouatta discolor*, *Alouatta macconnelli*
$Val^{3}Pro^{8}OT$	12	16	*Leontocebus fuscicollis*, *Leontocebus nigricollis*, *Saguinus bicolor*, *Saguinus geoffroyi*, *Saguinus imperator*, *Saguinus inustus*, *Saguinus labiatus*, *Saguinus martinsi*, *Saguinus midas*, *Saguinus mystax*, *Saguinus niger*, *Saguinus oedipus*
$Val^{3}OT$	3	3	*Alouatta juara*, *Alouatta seniculus*, *Alouatta seniculus puruensis*

Continued on next page

Table S3. Continued.

Variant	Species_Count	Samples_Count	Species_List
Pro^{8}OT	43	58	*Aotus azarae*, *Aotus griseimembra*, *Aotus nancymaae*, *Aotus nigriceps*, *Aotus trivirgatus*, *Aotus vociferans*, *Ateles belzebuth*, *Ateles chamek*, *Ateles geoffroyi*, *Ateles marginatus*, *Ateles paniscus*, *Callimico goeldii*, *Callithrix aurita*, *Callithrix geoffroyi*, *Callithrix jacchus*, *Callithrix kuhlii*, *Callithrix penicillata*, *Cebuella niveiventris*, *Cebuella pygmaea*, *Cebus albifrons*, *Cebus capucinus*, *Cebus imitator*, *Cebus olivaceus*, *Leontopithecus chrysomelas*, *Leontopithecus chrysopygas*, *Leontopithecus rosalia*, *Mico argentatus*, *Mico chrysoleucus*, *Mico humeralifer*, *Mico humilis*, *Mico mauesi*, *Mico saterei*, *Saimiri boliviensis*, *Saimiri cassiquiarensis*, *Saimiri macrodon*, *Saimiri oerstedii*, *Saimiri sciureus*, *Saimiri ustus*, *Sapajus apella*, *Sapajus apella macrocephalus*, *Sapajus libidinosus*, *Sapajus robustus*, *Sapajus xanthosternos*
Ala^{8}OT	8	10	*Cacajao ayresi*, *Cacajao calvus*, *Cacajao hosomi*, *Cacajao melanocephalus*, *Chiropotes chiropotes*, *Chiropotes sagulatus*, *Chiropotes satanas*, *Chiropotes utahickae*
Thr^{8}OT	9	9	*Chiropotes albinasus*, *Pithecia albicans*, *Pithecia chrysocephala*, *Pithecia hirsuta*, *Pithecia mittermeier*, *Pithecia mittermeieri*, *Pithecia pissinatti*, *Pithecia pithecia*, *Pithecia vanzolinii*

Supplementary Table 4. Overview of species included in each analytical stage, including OT variation, OTR variation, *OXT* PAML, *OXTR* PAML, explanatory-capacity analyses, and broad-set association analyses.

Species	OT_Var	OTR_Var	*OXT*_PAML	*OXTR*_PAML	Expl_Capacity	Broad_Set_Assoc
Total species	249	238	224	204	200	187
Allenopithecus nigroviridis	X	X	X	X	X	X
Allochrocebus lhoesti	X	X	X	X	X	X
Allochrocebus preussi	X	X	X	X	X	X
Allochrocebus solatus	X	X	X	X	X	X
Alouatta belzebul	X	X	X	X	X	X
Alouatta caraya	X	X	X	X	X	X
Alouatta clamitans	X	–	X	–	–	–
Alouatta discolor	X	X	X	X	–	–
Alouatta guariba	X	–	–	–	–	–
Alouatta juara	X	X	X	X	–	–
Alouatta macconnelli	X	X	X	X	X	X
Alouatta palliata	X	X	X	X	X	X
Alouatta seniculus	X	X	–	–	–	–
Alouatta seniculus puruensis	X	X	X	X	X	X
Alouatta ululata	X	–	–	–	–	–
Aotus azarae	X	X	X	X	X	X
Aotus griseimembra	X	X	X	X	X	X
Aotus lemurinus	–	X	–	X	–	–
Aotus nancymaae	X	X	X	X	X	X
Aotus nigriceps	X	–	X	–	X	X
Aotus trivirgatus	X	X	X	X	X	X
Aotus vociferans	X	X	X	X	X	X
Arctocebus calabarensis	–	X	–	X	–	–
Ateles belzebuth	X	X	X	X	X	X
Ateles chamek	X	X	X	X	X	X
Ateles geoffroyi	X	X	X	X	X	X
Ateles marginatus	X	X	X	X	X	X
Ateles paniscus	X	X	X	X	X	X
Avahi laniger	X	X	X	X	X	X
Avahi peyrierasi	X	X	X	X	X	X
Brachyteles arachnoides	X	–	X	–	–	–
Brachyteles hypoxanthus	X	X	X	X	X	X
Cacajao ayresi	X	X	X	X	X	X
Cacajao calvus	X	X	X	X	X	X
Cacajao hosomi	X	X	X	X	X	X
Cacajao melanocephalus	X	X	X	X	X	X
Callicebus personatus	X	–	X	–	–	–
Callimico goeldii	X	X	X	X	X	X
Callithrix aurita	X	–	X	–	–	–
Callithrix geoffroyi	X	X	X	X	X	X
Callithrix jacchus	X	X	X	X	X	X
Callithrix kuhlii	X	X	X	X	X	X
Callithrix penicillata	X	X	X	X	X	X

Continued on next page

Table S4. Continued.

Species	OT_Var	OTR_Var	*OXT*_PAML	*OXTR*_PAML	Expl_Capacity	Broad_Set_Assoc
Cebuella niveiventris	X	X	X	X	–	–
Cebuella pygmaea	X	X	X	X	X	X
Cebus albifrons	X	X	X	X	X	X
Cebus capucinus	X	X	X	X	X	X
Cebus imitator	X	X	–	–	–	–
Cebus olivaceus	X	X	X	X	X	X
Cebus unicolor	–	X	–	X	–	–
Cephalopachus bancanus	X	X	X	X	X	X
Cercocebus atys	X	X	–	–	–	–
Cercocebus atys lunulatus	X	X	X	X	X	X
Cercocebus chrysogaster	X	X	X	X	X	X
Cercocebus torquatus	X	X	X	X	X	X
Cercopithecus ascanius	X	X	X	X	X	X
Cercopithecus campbelli lowei	X	X	X	X	X	X
Cercopithecus cephus	X	X	–	–	–	–
Cercopithecus diana	X	X	X	X	X	X
Cercopithecus hamlyni	X	X	X	X	X	X
Cercopithecus neglectus	X	X	X	X	X	X
Cercopithecus nictitans	X	X	X	X	X	X
Cercopithecus petaurista	X	X	X	X	X	X
Cercopithecus pogonias	X	X	–	–	–	–
Cercopithecus roloway	X	X	X	X	X	X
Cheirogaleus major	X	X	X	X	X	X
Cheirogaleus medius	X	X	X	X	X	X
Cheracebus lucifer	X	X	X	X	X	X
Cheracebus lugens	X	X	X	X	X	X
Cheracebus regulus	X	X	X	X	X	X
Cheracebus torquatus	X	X	X	–	X	X
Chiropotes albinasus	X	X	X	X	X	X
Chiropotes chiropotes	X	X	X	X	X	X
Chiropotes sagulatus	X	X	–	–	–	–
Chiropotes satanas	X	–	X	–	–	–
Chiropotes utahickae	X	X	X	X	X	X
Chlorocebus pygerythrus	X	X	X	X	X	X
Chlorocebus sabaeus	X	X	X	X	X	X
Colobus angolensis	–	X	–	X	–	–
Colobus polykomos	X	X	X	X	X	X
Daubentonia madagascariensis	X	X	X	X	X	X
Erythrocebus patas	X	X	X	X	X	X
Eulemur albifrons	X	X	X	X	X	–
Eulemur coronatus	X	X	X	X	X	–
Eulemur flavifrons	X	X	X	X	X	–
Eulemur fulvus	X	X	–	–	–	–
Eulemur fulvus collaris	X	X	X	–	–	–
Eulemur macaco	X	X	X	X	X	–
Eulemur mongoz	X	X	X	X	X	X
Eulemur rubriventer	X	X	X	X	X	–
Eulemur rufus	X	X	X	X	X	–
Eulemur sanfordi	X	X	X	X	X	–
Galago senegalensis	X	X	X	X	X	X
Galagoides demidoff	X	X	X	X	X	X
Gorilla beringei	X	X	X	X	X	X
Gorilla gorilla	X	X	X	X	X	X
Hapalemur griseus	X	X	–	–	–	–
Hapalemur griseus alaotrensis	X	X	X	–	–	–
Hapalemur meridionalis	X	X	–	–	–	–
Hapalemur occidentalis	X	X	X	X	X	–
Homo sapiens	X	X	X	X	X	X
Hylobates agilis	X	X	X	X	X	X
Hylobates klossii	X	X	X	X	X	X
Hylobates moloch	X	X	X	X	X	X
Hylobates muelleri	X	X	X	X	X	X
Hylobates muelleri abbotti	X	X	–	–	–	–
Indri indri	X	X	X	X	X	–
Lagothrix lagotricha	X	X	X	X	X	X
Lagothrix poeppigii	X	X	X	X	X	X
Lemur catta	X	X	X	X	X	–
Leontocebus fuscicollis	X	X	–	–	–	–
Leontocebus nigricollis	X	X	X	X	X	X
Leontopithecus chrysomelas	X	X	X	X	X	X
Leontopithecus chrysopygas	X	–	–	–	–	–

Continued on next page

Table S4. Continued.

Species	OT_Var	OTR_Var	*OXT*_PAML	*OXTR*_PAML	Expl_Capacity	Broad_Set_Assoc
Leontopithecus rosalia	X	X	X	X	X	X
Lepilemur ankaranensis	X	X	X	–	X	X
Lepilemur dorsalis	X	X	X	–	X	X
Lepilemur ruficaudatus	X	X	X	–	X	X
Lepilemur septentrionalis	X	X	X	–	X	X
Loris lydekkerianus	X	X	X	X	X	X
Macaca arctoides	X	X	X	X	X	X
Macaca cyclopis	X	X	X	X	X	X
Macaca fascicularis	X	X	X	X	X	X
Macaca fuscata	X	X	–	–	–	–
Macaca fuscata fuscata	X	X	X	X	X	X
Macaca leonina	X	X	X	X	X	X
Macaca maura	X	X	X	X	X	X
Macaca mulatta	X	X	X	X	X	X
Macaca nemestrina	X	X	X	X	X	X
Macaca nigra	X	X	X	X	X	X
Macaca radiata	X	X	X	X	X	X
Macaca siberu	X	X	X	X	X	X
Macaca thibetana	X	X	X	X	X	X
Macaca tonkeana	X	X	X	X	X	X
Mandrillus leucophaeus	X	X	X	X	X	X
Mandrillus sphinx	X	X	X	X	X	X
Mico argentatus	X	X	X	X	X	X
Mico chrysoleucus	X	–	–	–	–	–
Mico humeralifer	X	X	X	X	X	X
Mico humilis	X	X	X	X	X	X
Mico mauesi	X	–	X	–	–	–
Mico saterei	X	–	X	–	–	–
Microcebus griseorufus	X	X	X	–	X	X
Microcebus mittermeieri	X	X	X	–	X	X
Microcebus murinus	X	X	X	–	X	X
Microcebus ravelobensis	X	X	X	–	X	X
Microcebus tavaratra	X	X	X	–	X	X
Miopithecus ogouensis	X	X	X	X	X	X
Mirza coquereli	X	X	X	–	X	X
Mirza zaza	X	X	X	–	X	X
Nasalis larvatus	X	X	X	X	X	X
Nomascus annamensis	X	X	–	–	–	–
Nomascus concolor	X	X	X	X	X	X
Nomascus gabriellae	X	X	X	X	X	X
Nomascus leucogenys	X	X	X	X	X	X
Nomascus siki	X	X	X	X	X	X
Nycticebus coucang	X	X	X	X	X	X
Otolemur crassicaudatus	X	X	X	X	X	X
Otolemur garnettii	X	X	X	X	X	X
Pan paniscus	X	X	X	X	X	X
Pan troglodytes	X	X	X	X	X	X
Papio anubis	X	X	X	X	X	X
Papio cynocephalus	X	X	X	X	X	X
Papio kindae	X	X	X	X	–	–
Papio papio	X	X	X	X	X	X
Papio ursinus	X	X	X	X	X	X
Perodicticus potto	X	X	–	–	–	–
Perodicticus potto ibeanus	X	X	X	X	X	X
Piliocolobus badius	X	X	X	X	X	X
Piliocolobus gordonorum	X	X	X	X	X	X
Piliocolobus kirkii	X	X	X	X	X	X
Piliocolobus tephrosceles	X	X	X	X	X	X
Pithecia albicans	X	X	X	X	X	X
Pithecia chrysocephala	X	X	X	X	–	–
Pithecia hirsuta	X	X	X	X	–	–
Pithecia mittermeier	X	–	–	–	–	–
Pithecia mittermeieri	X	X	X	X	–	–
Pithecia pissinatti	X	X	X	X	–	–
Pithecia pithecia	X	X	X	X	X	X
Pithecia vanzolinii	X	X	X	X	–	–
Plecturocebus bernhardi	X	X	X	X	X	X
Plecturocebus brunneus	X	X	X	X	X	X
Plecturocebus caligatus	X	X	X	X	X	X
Plecturocebus caquetensis	X	–	–	–	–	–
Plecturocebus cinerascens	X	X	X	X	X	X

Continued on next page

Table S4. Continued.

Species	OT_Var	OTR_Var	*OXT*_PAML	*OXTR*_PAML	Expl_Capacity	Broad_Set_Assoc
Plecturocebus cupreus	X	X	X	X	X	X
Plecturocebus donacophilus	X	X	X	X	X	X
Plecturocebus dubius	X	X	X	X	X	X
Plecturocebus grovesi	X	X	X	X	–	–
Plecturocebus hoffmannsi	X	X	X	X	X	X
Plecturocebus miltoni	X	X	X	X	–	–
Plecturocebus moloch	X	X	X	X	X	X
Pongo abelii	X	X	X	X	X	X
Pongo pygmaeus	X	X	X	X	–	–
Presbytis comata	X	X	X	X	X	X
Presbytis melalophos mitrata	X	X	X	X	X	X
Prolemur simus	X	X	X	X	X	–
Propithecus coquereli	X	X	X	X	X	X
Propithecus deckenii coronatus	X	X	X	X	X	X
Propithecus diadema	X	X	X	X	X	–
Propithecus edwardsi	X	X	X	X	X	X
Propithecus perrieri	X	X	X	X	X	–
Propithecus tattersalli	X	X	X	X	X	X
Propithecus verreauxi	X	X	X	X	X	X
Pygathrix cinerea	X	X	X	X	X	X
Pygathrix nemaeus	X	X	X	X	X	X
Pygathrix nigripes	X	X	X	X	X	X
Rhinopithecus bieti	X	X	X	X	X	X
Rhinopithecus roxellana	X	X	X	X	X	X
Saguinus bicolor	X	X	X	X	X	X
Saguinus geoffroyi	X	X	X	X	X	X
Saguinus imperator	X	X	X	X	X	X
Saguinus inustus	X	X	X	X	X	X
Saguinus labiatus	X	X	X	X	X	X
Saguinus martinsi	X	X	X	X	X	X
Saguinus midas	X	X	X	X	X	X
Saguinus mystax	X	X	X	X	X	X
Saguinus niger	X	X	X	X	X	X
Saguinus oedipus	X	X	X	X	X	X
Saimiri boliviensis	X	X	X	X	X	X
Saimiri cassiquiarensis	X	X	X	X	–	–
Saimiri macrodon	X	X	X	X	X	X
Saimiri oerstedii	X	X	X	X	X	X
Saimiri sciureus	X	X	–	–	–	–
Saimiri ustus	X	X	X	X	X	X
Sapajus apella	X	X	–	–	–	–
Sapajus apella macrocephalus	X	X	X	X	X	X
Sapajus libidinosus	X	–	X	–	–	–
Sapajus robustus	X	X	X	X	X	X
Sapajus xanthosternos	X	X	X	X	X	X
Semnopithecus entellus	X	X	X	X	X	X
Semnopithecus hypoleucos	X	X	–	–	–	–
Semnopithecus priam	X	X	X	X	X	X
Semnopithecus schistaceus	X	X	–	–	–	–
Symphalangus syndactylus	X	X	X	X	X	X
Tarsius lariang	X	X	X	X	X	X
Tarsius wallacei	X	X	X	X	X	X
Theropithecus gelada	X	X	X	X	X	X
Trachypithecus auratus	X	X	X	X	X	X
Trachypithecus cristatus	X	X	X	X	X	X
Trachypithecus francoisi	X	X	X	X	X	X
Trachypithecus geei	X	X	X	X	X	X
Trachypithecus germaini	X	X	X	X	X	X
Trachypithecus hatinhensis	X	X	X	X	X	X
Trachypithecus johnii	X	X	X	X	X	X
Trachypithecus laotum	X	X	X	X	X	X
Trachypithecus obscurus	X	X	X	X	X	X
Trachypithecus phayrei	X	X	–	–	–	–
Trachypithecus phayrei crepuscula	X	X	X	–	–	–
Trachypithecus pileatus	X	X	X	X	X	X
Trachypithecus poliocephalus	X	X	X	X	X	X
Trachypithecus vetulus	X	X	X	X	X	X
Varecia rubra	X	X	X	X	X	X

Supplementary Table 5. Model comparison Likelihood Ratio Test (LRT) Under Site Models for *OXT* with intraspecific variation

Comparison	LRT Statistic ($2\Delta\ell$)	Degrees of Freedom	p-value
M1a vs. M2a	-5.65012999999999	2	1
M7 vs. M8	1.0886639999999943	2	0.58022924704959
M8a vs. M8	-0.32592399999998634	1	1

Supplementary Table 6. Log-likelihood Values and Parameter Estimates for the Site Models under different codon substitution models (NsSites)

Model	d_N/d_S	Estimated parameters	ℓ
M1a	0.129200	p_0 = 0.88835; p_1= 0.11165; ω_0 = 0.01976; ω_1 = 1	-181.719129
M2a	0.128900	p_0 = 0.88836; p_1= 0.11164; p_2= 0; ω_0 = 0.01939; ω_1 = 1; ω_2 = 5.98511	-184.544194
M7	0.096200	beta (p = 0.08770, q = 0.79734)	-177.844734
M8 β&ω	0.096200	p_0 = 0.99999; (p_1= 0.00001); p_2= 0.08770; q = 0.79735; ω_{11}= 1.00000	-177.300402
M8a	0.092200	p_0 = 0.99999; (p_1= 0.00001); p_2= 0.03250; q = 0.33376; ω_{11}= 1	-177.137440

Supplementary Table 7. Model comparison Likelihood Ratio Test (LRT) Under Site Models for *OXT* without intraspecific variation

Comparison	LRT Statistic ($2\Delta\ell$)	Degrees of Freedom	p-value
M1a vs. M2a	3.1156539999999495	2	0.21059319336684795
M7 vs. M8	6.829410000000024	2	0.03288610667216533
M8a vs. M8	-1.2446960000000047	1	1

Supplementary Table 8. Log-likelihood Values and Parameter Estimates for the Site Models under different codon substitution models (NsSites) for *OXT* without intraspecific variation

Model	d_N/d_S	Estimated parameters	ℓ
M1a	0.140000	p_0 = 0.87174; p_1= 0.12826; ω_0 = 0.01351; ω_1 = 1	-129.720702
M2a	0.250700	p_0 = 0.74930; p_1= 0.12283; p_2= 0.12787; ω_0 = 0.00000; ω_1 = 1; ω_2 = 1	-128.162875
M7	0.614700	beta (p = 0.03545, q = 0.02100)	-129.374906
M8 β&ω	0.115800	p_0 = 0.99999; (p_1= 0.00001); p_2= 0.01359; q = 0.08949; ω_{11}= 3.94446	-125.960201
M8a	0.140400	p_0 = 0.97207; (p_1= 0.02793); p_2= 0.01151; q = 0.07414; ω_{11}= 1	-125.337853

Supplementary Table 9. Model comparison Likelihood Ratio Test (LRT) Under Site Models for *OXT*R with intraspecific variation

Comparison	LRT Statistic ($2\Delta\ell$)	Degrees of Freedom	p-value
M1a vs. M2a	9.061745999999403	2	0.010771268649472527
M7 vs. M8	46.767165999997815	2	6.992604027088808e-11
M8a vs. M8	5.703381999999692	1	0.01693225601754586

Supplementary Table 10. Log-likelihood Values and Parameter Estimates for the Site Models under different codon substitution models (NsSites) for *OXT*R with intraspecific variation

Model	d_N/d_S	Estimated parameters	ℓ
M1a	0.117800	p_0 = 0.91368; p_1= 0.08632ω_0 = 0.03444; ω_1 = 1	-11108.515892
M2a	0.131800	p_0 = 0.91249; p_1= 0.08184; p_2= 0.00568; ω_0 = 0.03464; ω_1 = 1; ω_2 = 3.23372	-11103.985019
M7	0.088800	beta (p = 0.15849, q = 1.53817)	-10993.222510
M8 β&ω	0.084400	p_0 = 0.99131; (p_1= 0.00869); p_2= 0.18340; q = 2.40521; ω_{11}= 2.19070	-10969.838927
M8a	0.071800	p_0 = 0.98200; (p_1= 0.01800); p_2= 0.19721; q = 3.12644; ω_{11}= 1	-10972.690618

Supplementary Table 11. Model comparison Likelihood Ratio Test (LRT) Under Site Models for *OXT*R without intraspecific variation

Comparison	LRT Statistic ($2\Delta\ell$)	Degrees of Freedom	p-value
M1a vs. M2a	0.7498279999999795	2	0.6873483882106239
M7 vs. M8	0.004309999998440617	2	1
M8a vs. M8	-0.39418799999839393	1	1

Supplementary Table 12. Log-likelihood Values and Parameter Estimates for the Site Models under different codon substitution models (NsSites) for *OXT*R without intraspecific variation

Model	d_N/d_S	Estimated parameters	ℓ
M1a	0.092500	p_0 = 0.94070; p_1= 0.05930; ω_0 = 0.03531; ω_1 = 1	-4558.981742
M2a	0.092400	p_0 = 0.94079; p_1= 0.02368; p_2= 0.03553; ω_0 = 0.03526; ω_1 = 1; ω_2 = 1	-4558.606828
M7	0.062500	beta (p = 0.16775, q = 2.31787)	-4530.022047
M8 β&ω	0.062500	p_0 = 0.99999; (p_1= 0.00001); p_2= 0.16776; q = 2.31805; ω_{11}= 1.34215	-4530.024202
M8a	0.062500	p_0 = 0.99999; (p_1= 0.00001); p_2= 0.16774; q = 2.31793; ω_{11}= 1	-4529.827108

Supplementary Table 13. Branch-site Model Analysis for the *OXT* Gene With Intraspecific Variation

Model	Estimated parameters	Log-likelihood ($\ln L$)	LRT Statistic ($2\Delta\ell$)	p-value (χ^2_1 50:50 mixture)
Alternative	p_0 = 0.65665; p_1 = 0.18834; p_{2a} = 0.12046; p_{2b} = 0.03455; ω_0 (background) = 0.00000; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000	-177.224962	0 (original -3.739948)	1
Null	p_0 = 0.68402; p_1 = 0.08602; p_{2a} = 0.20428; p_{2b} = 0.02569; ω_0 (background) = 0.00602; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000 (fixed)	-175.354988	–	–

Supplementary Table 14. Branch-site Model Analysis for the *OXT* Gene Without Intraspecific Variation

Model	Estimated parameters	Log-likelihood ($\ln L$)	LRT Statistic ($2\Delta\ell$)	p-value (χ^2_1 50:50 mixture)
Alternative	p_0 = 0.74090; p_1 = 0.10656; p_{2a} = 0.13335; p_{2b} = 0.01918; ω_0 (background) = 0.00430; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000	-126.573171	0 (original -0.087168)	1
Null	p_0 = 0.73716; p_1 = 0.10589; p_{2a} = 0.13723; p_{2b} = 0.01971; ω_0 (background) = 0.00418; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000 (fixed)	-126.529587	–	–

Supplementary Table 15. Branch-Site Model Analysis for the *OXTR* Gene With Intraspecific Variation

Model	Estimated parameters	Log-likelihood ($\ln L$)	LRT Statistic ($2\Delta\ell$)	p-value (χ^2_1 50:50 mixture)
Alternative	p_0 = 0.88807; p_1 = 0.06605; p_{2a} = 0.04270; p_{2b} = 0.00318; ω_0 (background) = 0.03085; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000	-11074.947806	0 (original -0.00017)	1
Null	p_0 = 0.88807; p_1 = 0.06609; p_{2a} = 0.04266; p_{2b} = 0.00318; ω_0 (background) = 0.03085; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000 (fixed)	-11074.947891	–	–

Supplementary Table 16. Branch-Site Model Analysis for the *OXTR* Gene Without Intraspecific Variation

Model	Estimated parameters	Log-likelihood ($\ln L$)	LRT Statistic ($2\Delta\ell$)	p-value (χ^2_1 50:50 mixture)
Alternative	p_0 = 0.88470; p_1 = 0.04509; p_{2a} = 0.06680; p_{2b} = 0.00340; ω_0 (background) = 0.02955; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000	-4548.100554	0 (original -0.506306)	1
Null	p_0 = 0.89594; p_1 = 0.04567; p_{2a} = 0.05556; p_{2b} = 0.00283; ω_0 (background) = 0.03083; ω_1 (neutral) = 1.00000; ω_2 (foreground) = 1.00000 (fixed)	-4548.353707	–	–

Supplementary Table 17. Top 39 informative features identified by the BroadSet_RF approach as sufficient to classify 187 primate species according to the social monogamy phenotype.

Feature	Importance
R362_A	0.062947
R182_D	0.044558
R182_E	0.043553
R362_T	0.038136
R16_A	0.037545
O8_A	0.034163
R16_S	0.031149
R162_L-R162_F	0.029117
R345_M-R345_L	0.027720
R350_R	0.026810
O8_P	0.024500
R355_N	0.022163
R218_A	0.021012
R5_F	0.019684
R172_V	0.019136
R350_S	0.019109
R218_T	0.018684
R355_K	0.018369
R258_G-R258_A	0.018122
R172_M	0.018012
R248_V	0.017908
R360_T-R360_M	0.017550
O3_V-O3_I	0.015536
R286_V-R286_A+R312_T	0.013614
O8_L	0.013313
R169_V-R169_A	0.013112
R33_R-R33_Q	0.012765
R236_V-R236_A	0.011622
R14_V	0.010920
R248_A	0.010670
R303_N	0.010565
R5_L	0.010499
R14_I	0.010456
R312_I	0.010420
R247_S	0.008766
R303_D	0.008656
R66_T+R343_R-R66_I+R343_H	0.008498
R141_V-R141_I	0.008368
R178_R	0.008145

Supplementary Table 18. Features significantly associated with the social monogamy phenotype according to Fisher's exact test after Bonferroni correction.

Feature	p-value	Adjusted p-value (Bonferroni)
R362_A	6.725e-15	9.550e-13
R362_T	1.365e-13	1.939e-11
R16_S	3.160e-12	4.487e-10
R16_A	4.822e-12	6.847e-10
R172_M	1.279e-09	1.816e-07
R5_F	2.592e-09	3.681e-07
R172_V	3.444e-09	4.890e-07
R248_V	5.783e-09	8.212e-07
R169_V-R169_A	4.054e-08	5.757e-06
R248_A	1.371e-07	1.946e-05
O8_P	6.171e-07	8.763e-05
R5_L	1.985e-06	2.819e-04
O8_L	1.056e-05	0.001
O3_V-O3_I	2.044e-05	0.003
R258_G-R258_A	2.736e-05	0.004
R33_R-R33_Q	3.182e-05	0.005
R182_D	1.358e-04	0.019

Supplementary Table 19. Top 10 features ranked in the phylogenetic logistic regression analysis (PhyloLogReg) of the social monogamy phenotype across primates.

Feature	α	β_1	p-value	Adjusted p-value (Bonferroni)
R362_T	0.022831	-4.110691	0.167	1.000
R201_V-R201_I	0.024594	-2.472281	0.195	1.000
R5_del+R7_del+R55_G+R279_V-R10_S+R55_S+R279_I	0.024432	-2.515533	0.215	1.000
R353_K	0.025527	2.109753	0.257	1.000
R106_Y-R106_H	0.016928	-1.889569	0.351	1.000
R286_V-R286_A+R312_T	0.015570	-1.901275	0.369	1.000
R355_K	0.023699	-2.053878	0.379	1.000
R242_Q	0.015074	1.786777	0.407	1.000
O8_T	0.013139	0.886388	0.472	1.000
R51_L+R355_R	0.017972	1.500182	0.494	1.000

Supplementary Table 20. Features with significant importance in Random Forest permutation tests (RF_PermImp) after Bonferroni correction.

Feature	p-value	Importance	Adjusted p-value (Bonferroni)
R172_V	0	0.016292	0
R169_V-R169_A	0	0.017084	0
R16_A	0	0.041141	0
R182_D	0	0.053052	0
R172_M	0	0.011533	0
R218_T	0	0.020680	0
R362_T	0	0.042110	0
R16_S	0	0.035880	0
R218_A	0	0.016910	0
R362_A	0	0.055009	0
R5_F	0	0.022352	0

Supplementary Table 21. Pathogenicity predictions for missense variants corresponding to the selected features using SIFT, PolyPhen-2, and AlphaMissense.

Gene	protein_variant	Feature	SIFT	PolyPhen-2	AlphaMissense
OXT	p.Ile22Val	O3_V-O3_I	deleterious (0.01)	possibly_damaging (0.749)	likely_pathogenic (0.5996)
OXT	p.Leu27Ala	O8_L, O8_A	tolerated (0.19)	benign (0.103)	likely_benign (0.1516)
OXT	p.Leu27Pro	O8_L, O8_P	tolerated (0.12)	benign (0)	likely_benign (0.0702)
OXTR	p.Leu5Phe	R5_F, R5_L	tolerated (0.31)	unknown (0)	likely_benign (0.0852)
OXTR	p.Ala14Val	R14_V	tolerated (0.75)	unknown (0)	likely_benign (0.0916)
OXTR	p.Ala14Ile	R14_I	tolerated (0.48)	unknown (0)	likely_benign (0.1549)
OXTR	p.Ala16Thr	R16_A	tolerated (0.37)	unknown (0)	likely_benign (0.0836)
OXTR	p.Ala16Ser	R16_A, R16_S	tolerated (1)	unknown (0)	likely_benign (0.0775)
OXTR	p.Ala16Cys	R16_A	tolerated (0.13)	unknown (0)	likely_benign (0.2754)
OXTR	p.Ala16Gly	R16_A	tolerated (0.83)	unknown (0)	likely_benign (0.0594)
OXTR	p.Arg33Gly	R33_R-R33_Q	tolerated (0.57)	benign (0)	likely_benign (0.0763)
OXTR	p.Thr66Ile	R66_T+R343_R-R66_I+R343_H	tolerated (0.12)	benign (0.038)	likely_benign (0.1514)
OXTR	p.Ile141Val	R141_V-R141_I	tolerated (0.4)	benign (0.145)	likely_benign (0.1224)
OXTR	p.Leu162Phe	R162_L-R162_F	tolerated (0.06)	probably_damaging (0.927)	likely_benign (0.1095)
OXTR	p.Ala169Val	R169_V-R169_A	tolerated (0.68)	benign (0.006)	likely_benign (0.0627)
OXTR	p.Val172Ala	R172_V	tolerated (0.32)	benign (0.035)	likely_benign (0.1282)
OXTR	p.Val172Met	R172_V,R172_M	tolerated (0.19)	benign (0.101)	likely_benign (0.1622)
OXTR	p.Arg178Gly	R178_Q	deleterious (0.01)	benign (0.058)	ambiguous (0.3586)
OXTR	p.Asp182Ala	R182_D	tolerated (0.07)	benign (0.027)	likely_benign (0.0932)
OXTR	p.Asp182Asn	R182_D	tolerated (1)	benign (0.003)	likely_benign (0.0894)
OXTR	p.Asp182Glu	R182_D,R182_E	missing	missing	likely_benign (0.0904)
OXTR	p.Ala218Thr	R218_A, R218_T	tolerated (0.66)	benign (0.011)	likely_benign (0.0838)
OXTR	p.Ala236Val	R236_V-R236_A	tolerated (0.14)	benign (0.038)	likely_benign (0.0998)
OXTR	p.Ala247Ser	R247_S	tolerated (0.66)	benign (0.012)	likely_benign (0.0871)
OXTR	p.Ala248Val	R248_V, R248_A	tolerated (0.1)	benign (0.017)	likely_benign (0.0992)
OXTR	p.Ala258Gly	R258_G-R258_A	deleterious (0.04)	benign (0.318)	likely_benign (0.0946)
OXTR	p.Val286Ala	R286_V-R286_A+R312_T	tolerated (0.21)	benign (0.151)	likely_benign (0.1332)
OXTR	p.Asn303Glu	R303_N	missing	missing	likely_benign (0.1256)
OXTR	p.Asn303Asp	R303_N, R303_D	tolerated (0.31)	benign (0.014)	likely_benign (0.0683)
OXTR	p.Ile312Thr	R286_V-R286_A+R312_T	tolerated (1)	benign (0.021)	ambiguous (0.4424)
OXTR	p.Arg343His	R66_T+R343_R-R66_I+R343_H	tolerated (0.55)	benign (0.01)	likely_benign (0.0805)
OXTR	p.Leu345Met	R345_M-R345_L	tolerated (0.22)	possibly_damaging (0.625)	likely_benign (0.1435)

Continued on next page

Gene	protein_variant	Feature	SIFT	PolyPhen-2	AlphaMissense
OXTR	p.Ser350Arg	R350_R, R350_S	missing	missing	likely_benign (0.0961)
OXTR	p.Ser350Asn	R350_S	tolerated (0.62)	benign (0.005)	likely_benign (0.0828)
OXTR	p.Arg355Lys	R355_K	tolerated (0.37)	benign (0.001)	likely_benign (0.0825)
OXTR	p.Arg355Asn	R355_N	missing	missing	likely_benign (0.1419)
OXTR	p.Thr360Met	R360_T-R360_M	tolerated (0.13)	probably_damaging (0.954)	likely_benign (0.0832)
OXTR	p.Ala362Leu	R362_A	tolerated (0.66)	unknown (0)	likely_benign (0.0746)
OXTR	p.Ala362Phe	R362_A	tolerated (0.45)	unknown (0)	likely_benign (0.1091)
OXTR	p.Ala362Thr	R362_A, R362_T	tolerated (0.74)	unknown (0)	likely_benign (0.0636)
OXTR	p.Ala362Val	R362_A	tolerated (0.63)	unknown (0)	likely_benign (0.0648)
OXTR	p.Ala362Cys	R362_A	tolerated (0.22)	unknown (0)	likely_benign (0.132)

Supplementary Table 22. DDMut predictions of protein stability changes for single and double substitutions.

mutations	average_distance	single_predictions	sum_single_predictions	ddmut_prediction	prediction_reverse
A A169V; A V172M	5.51	-0.01; -0.30	-0.31	1.20	-0.59

Legends for Datasets S1 to S7

Dataset S1 (D1_Feature_Importance.xlsx)
Feature importance scores for all 143 genetic features calculated using Random Forest (1,000 trees), ranked by mean decrease in impurity.

Dataset S2 (D2_Fisher_exact.xlsx)
Results of Fisher's exact test for all 143 features, including contingency table counts, raw p-values, and Bonferroni-corrected p-values.

Dataset S3 (D3_Phyloglm.xlsx)
Results of phylogenetic logistic regression (PhyloLogReg) for all 143 features, including coefficients (β_1), α, and adjusted p-values.

Dataset S4 (D4_RF_PermImp.xlsx)
Random Forest permutation importance results (RF_PermImp) for all 143 features, including observed importance, empirical p-values, and Bonferroni-corrected p-values.

Dataset S5 (D5_Combination_select.xlsx)
Genotype combinations identified across amino acid positions O3, O8, and R33 in Platyrrhini species with social monogamy phenotype data.

Dataset S6 (D6_Combination_functional.xlsx)
Genotype combinations identified across amino acid positions O3, O8, R33, R141, R343, and R345 in Platyrrhini species with social monogamy phenotype data.

Dataset S7 (D7_Combination_Platyrrhini_all.xlsx)
Genotype combinations identified across all variable amino acid positions retained from the BroadSet_RF approach in Platyrrhini species with social monogamy phenotype data.

Supplementary References

1. P Vargas-Pinilla, et al., From molecular variations to behavioral adaptations: Unveiling adaptive epistasis in primate oxytocin system. *Am. J. Biol. Anthropol.* p. e24947 (2024).
2. FJ Martin, et al., Ensembl 2023. *Nucleic acids research* **51**, D933–D941 (2023).
3. NA O'Leary, et al., Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic acids research* **44**, D733–D745 (2016).
4. LF Kuderna, et al., A global catalog of whole-genome diversity from 233 primate species. *Science* **380**, 906–913 (2023).
5. RC Edgar, Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).
6. A Larsson, Aliview: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
7. A Criscuolo, S Gribaldo, Bmge (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC evolutionary biology* **10**, 1–21 (2010).

8. A Estrada, et al., Impending extinction crisis of the world's primates: Why primates matter. *Sci. advances* **3**, e1600946 (2017).
9. Y Shao, et al., Phylogenomic analyses provide insights into primate evolution. *Science* **380**, 913–924 (2023).
10. AB Rylands, RA Mittermeier, The diversity of the new world primates (platyrrhini): an annotated taxonomy in *South American primates: comparative perspectives in the study of behavior, ecology, and conservation*. (Springer), pp. 23–54 (2009).
11. H Byrne, et al., Phylogenetic relationships of the new world titi monkeys (callicebus): first appraisal of taxonomy based on molecular evidence. *Front. zoology* **13**, 1–26 (2016).
12. JC Buckner, JWL Alfaro, AB Rylands, ME Alfaro, Biogeography of the marmosets and tamarins (callitrichidae). *Mol. phylogenetics evolution* **82**, 413–425 (2015).
13. L Pozzi, et al., Remarkable ancient divergences amongst neglected lorisiform primates. *Zool. J. Linnean Soc.* **175**, 661–674 (2015).
14. Z Yang, Paml 4: phylogenetic analysis by maximum likelihood. *Mol. biology evolution* **24**, 1586–1591 (2007).
15. LST Ho, et al., Package 'phylolm'. *See http://cran. r-project. org/web/packages/phylolm/index. html (accessed Febr. 2018)* (2016).
16. S Kumar, et al., Timetree 5: an expanded resource for species divergence times. *Mol. biology evolution* **39**, msac174 (2022).
17. G Van Rossum, FL Drake Jr, *Python tutorial*. (Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands) Vol. 620, (1995).
18. DG Kleiman, Monogamy in mammals. *The Q. review biology* **52**, 39–69 (1977).
19. T Clutton-Brock, Social evolution in mammals. *Science* **373**, eabc9699 (2021).
20. D Lukas, TH Clutton-Brock, The evolution of social monogamy in mammals. *Science* **341**, 526–530 (2013).
21. JA French, J Cavanaugh, AC Mustoe, SB Carp, SL Womack, Social monogamy in nonhuman primates: phylogeny, phenotype, and physiology. *The J. Sex Res.* **55**, 410–434 (2018).
22. E Fernandez-Duque, M Huck, S Van Belle, A Di Fiore, The evolution of pair-living, sexual monogamy, and cooperative infant care: Insights from research on wild owl monkeys, titis, sakis, and tamarins. *Am. J. Phys. Anthropol.* **171**, 118–173 (2020).
23. M Huck, A Di Fiore, E Fernandez-Duque, Of apples and oranges? the evolution of "monogamy" in non-human primates. *Front. Ecol. Evol.* **7**, 472 (2020).
24. S Shultz, C Opie, QD Atkinson, Stepwise evolution of stable sociality in primates. *Nature* **479**, 219–222 (2011).
25. PM Kappeler, L Pozzi, Evolutionary transitions toward pair living in nonhuman primates as stepping stones toward more complex societies. *Sci. advances* **5**, eaay1276 (2019).
26. CA Olivier, et al., Primate social organization evolved from a flexible pair-living ancestor. *Proc. Natl. Acad. Sci.* **121**, e2215401120 (2024).
27. C Opie, QD Atkinson, RI Dunbar, S Shultz, Male infanticide leads to social monogamy in primates. *Proc. Natl. Acad. Sci.* **110**, 13328–13332 (2013).
28. W McKinney, , et al., pandas: a foundational python library for data analysis and statistics. *Python for high performance scientific computing* **14**, 1–9 (2011).
29. CR Harris, et al., Array programming with numpy. *nature* **585**, 357–362 (2020).
30. Y Waltenspühl, et al., Structural basis for the activation and ligand recognition of the human oxytocin receptor. *Nat. Commun.* **13**, 4153 (2022).
31. F Musiani, S Al Sadi, A Giorgetti, M Alfonso-Prieto, P Carloni, Structural and functional role of the magnesium ion in the human oxytocin receptor. *J. Inorg. Biochem.* p. 113309 (2026).
32. Y Waltenspühl, J Schöppe, J Ehrenmann, L Kummer, A Plückthun, Crystal structure of the human oxytocin receptor. *Sci. Adv.* **6**, eabb5419 (2020).
33. T UniProt Consortium, Uniprot: the universal protein knowledgebase. *Nucleic acids research* **46**, 2699–2699 (2018).
34. VJ Wesley, SR Hawtin, HC Howard, M Wheatley, Agonist-specific, high-affinity binding epitopes are contributed by an arginine in the n-terminus of the human oxytocin receptor. *Biochemistry* **41**, 5086–5092 (2002).
35. J Koehbach, T Stockner, C Bergmayr, M Muttenthaler, CW Gruber, Insights into the molecular evolution of oxytocin receptor ligand binding. *Biochem. Soc. Transactions* **41**, 197–204 (2013).
36. JG Meyerowitz, et al., The oxytocin signaling complex reveals a molecular switch for cation dependence. *Nat. Struct. & Mol. Biol.* **29**, 274–281 (2022).
37. G Gimpl, F Fahrenholz, The oxytocin receptor system: structure, function, and regulation. *Physiol. reviews* **81**, 629–683 (2001).
38. G Pándy-Szekeres, et al., Gpcrdb in 2018: adding gpcr structure models and ligands. *Nucleic acids research* **46**, D440–D446 (2018).
39. Z Yang, JP Bielawski, Statistical methods for detecting molecular adaptation. *Trends ecology & evolution* **15**, 496–503 (2000).
40. S Kryazhimskiy, JB Plotkin, The population genetics of dn/ds. *PLoS genetics* **4**, e1000304 (2008).
41. S Álvarez-Carretero, P Kapli, Z Yang, Beginner's guide on the use of paml to detect positive selection. *Mol. biology evolution* **40**, msad041 (2023).
42. P Virtanen, et al., Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. methods* **17**, 261–272 (2020).

43. WS Wong, Z Yang, N Goldman, R Nielsen, Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041–1051 (2004).
44. Z Yang, WS Wong, R Nielsen, Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. biology evolution* **22**, 1107–1118 (2005).
45. WJ Swanson, R Nielsen, Q Yang, Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. biology evolution* **20**, 18–20 (2003).
46. Z Yang, R Nielsen, Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. molecular evolution* **46**, 409–418 (1998).
47. P Vargas-Pinilla, et al., Evolutionary pattern in the oxt-oxtr system in primates: coevolution and positive selection footprints. *Proc. Natl. Acad. Sci.* **112**, 88–93 (2015).
48. C Meyer, et al., Understanding the causes of errors in eukaryotic protein-coding gene prediction: a case study of primate proteomes. *BMC bioinformatics* **21**, 513 (2020).
49. G Slodkowicz, N Goldman, Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. *Proc. Natl. Acad. Sci.* **117**, 5977–5986 (2020).
50. J Zhang, R Nielsen, Z Yang, Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. biology evolution* **22**, 2472–2479 (2005).
51. F Pedregosa, et al., Scikit-learn: Machine learning in python. *J. machine Learn. research* **12**, 2825–2830 (2011).
52. L Breiman, J Friedman, RA Olshen, CJ Stone, *Classification and regression trees*. (Routledge), (2017).
53. L Ein-Dor, I Kela, G Getz, D Givol, E Domany, Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* **21**, 171–178 (2005).
54. M Zucknick, S Richardson, EA Stronach, Comparing the characteristics of gene expression profiles derived by univariate and multivariate classification methods. *Stat. applications genetics molecular biology* **7** (2008).
55. Y Drier, E Domany, Do two machine-learning based prognostic signatures for breast cancer capture the same biological processes? *PloS one* **6**, e17795 (2011).
56. Z He, W Yu, Stable feature selection for biomarker discovery. *Comput. biology chemistry* **34**, 215–225 (2010).
57. JD Hunter, Matplotlib: A 2d graphics environment. *Comput. science & engineering* **9**, 90–95 (2007).
58. ML Waskom, Seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
59. RA Fisher, On the interpretation of χ 2 from contingency tables, and the calculation of p. *J. royal statistical society* **85**, 87–94 (1922).
60. OJ Dunn, Multiple comparisons among means. *J. Am. statistical association* **56**, 52–64 (1961).
61. S Seabold, J Perktold, Statsmodels: econometric and statistical modeling with python. *SciPy* **7** (2010).
62. AR Ives, T Garland Jr, Phylogenetic logistic regression for binary dependent variables. *Syst. biology* **59**, 9–26 (2010).
63. J Felsenstein, Phylogenies and the comparative method. *The Am. Nat.* **125**, 1–15 (1985).
64. A Grafen, The phylogenetic regression. *Philos. Transactions Royal Soc. London. B, Biol. Sci.* **326**, 119–157 (1989).
65. EP Martins, T Garland Jr, Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* **45**, 534–557 (1991).
66. R Diaz-Uriarte, T Garland Jr, Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from brownian motion. *Syst. Biol.* **45**, 27–47 (1996).
67. EP Martins, JA Diniz-Filho, EA Housworth, Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution* **56**, 1–13 (2002).
68. M Pagel, Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. Royal Soc. London. Ser. B: Biol. Sci.* **255**, 37–45 (1994).
69. GB Cybis, et al., Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. *The annals applied statistics* **9**, 969 (2015).
70. RC Team, Ra language and environment for statistical computing, r foundation for statistical. *Computing* (2020).
71. E Paradis, K Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r. *Bioinformatics* **35**, 526–528 (2019).
72. VA Huynh-Thu, Y Saeys, L Wehenkel, P Geurts, Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. *Bioinformatics* **28**, 1766–1774 (2012).
73. T Kolisnik, F Keshavarz-Rahaghi, RV Purcell, AN Smith, OK Silander, pyrforest: A comprehensive r package for genomic data analysis featuring scikit-learn random forests in r. *Briefings Funct. Genomics* p. elae038 (2024).
74. R Grantham, Amino acid difference formula to help explain protein evolution. *science* **185**, 862–864 (1974).
75. P Kumar, S Henikoff, PC Ng, Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. *Nat. protocols* **4**, 1073–1081 (2009).
76. IA Adzhubei, et al., A method and server for predicting damaging missense mutations. *Nat. methods* **7**, 248–249 (2010).
77. J Cheng, et al., Accurate proteome-wide missense variant effect prediction with alphamissense. *Science* **381**, eadg7492 (2023).
78. AD McLachlan, Tests for comparing related amino-acid sequences. cytochrome c and cytochrome c551. *J. molecular biology* **61**, 409–424 (1971).
79. WH Li, CI Wu, CC Luo, Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* **21**, 58–71 (1984).

80. Gt Ramachandran, V Sasisekharan, Conformation of polypeptides and proteins. *Adv. protein chemistry* **23**, 283–437 (1968).
81. Y Song, et al., High-resolution comparative modeling with rosettacm. *Structure* **21**, 1735–1742 (2013).
82. M Baek, et al., Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
83. Y Zhou, Q Pan, DE Pires, CH Rodrigues, DB Ascher, Ddmut: predicting effects of mutations on protein stability using deep learning. *Nucleic Acids Res.* **51**, W122–W128 (2023).
84. GK Hochberg, JW Thornton, Reconstructing ancient proteins to understand the causes of structure and function. *Annu. review biophysics* **46**, 247–269 (2017).
85. D Ren, et al., Genetic diversity in oxytocin ligands and receptors in new world monkeys. *PloS one* **10**, e0125775 (2015).
86. F Wiens, Ph.D. thesis (2002).
87. RM Nowak, *Walker's mammals of the world.* (JHU press) Vol. 1, (1999).
88. AF Richard, Preliminary observations on the birth and development of propithecus verreauxi to the age of six months. *Primates* **17**, 357–366 (1976).
89. ST Pochron, WT Tucker, PC Wright, Demography, life history, and social structure in propithecus diadema edwardsi from 1986–2000 in ranomafana national park, madagascar. *Am. J. Phys. Anthropol. The Off. Publ. Am. Assoc. Phys. Anthropol.* **125**, 61–72 (2004).
90. EJ Rode-Margono, KAI Nekaris, PM Kappeler, C Schwitzer, The largest relative testis size among primates and aseasonal reproduction in a nocturnal lemur, m irza zaza. *Am. J. Phys. Anthropol.* **158**, 165–169 (2015).
91. KJ Kling, PC Wright, Mating systems of lemurs (2018).
92. P Lahann, Habitat utilization of three sympatric cheirogaleid lemur species in a littoral rain forest of southeastern madagascar. *Int. J. Primatol.* **29**, 117–134 (2008).
93. LM Porter, Social organization, reproduction and rearing strategies of callimico goeldii: new clues from the wild. *Folia Primatol.* **72**, 69–79 (2001).
94. C Schradin, DM Reeder, SP Mendoza, G Anzenberger, Prolactin and paternal care: comparison of three species of monogamous new world monkeys (callicebus cupreus, callithrix jacchus, and callimico goeldii). *J. Comp. Psychol.* **117**, 166 (2003).
95. G Anzenberger, B Falk, Monogamy and family life in callitrichid monkeys: deviations, social dynamics and captive management. *Int. Zoo Yearb.* **46**, 109–122 (2012).
96. P Vargas-Pinilla, et al., Progesterone response element variation in the oxtr promoter region and paternal care in new world monkeys. *Behav. genetics* **47**, 77–87 (2017).
97. T Bartlett, The hylobatidae, small apes of asia. *Primates perspective* pp. 274–289 (2007).
98. GJ Fletcher, JA Simpson, L Campbell, NC Overall, Pair-bonding, romantic love, and evolution: The curious case of homo sapiens. *Perspectives on psychological science* **10**, 20–36 (2015).