Evolutionary History of Histamine Receptors: Early Vertebrate Origin and Expansion of the H_3-H_4 Subtypes

Infant Sagayaraj Ravhe[1], Arunkumar Krishnan[1,2], Narayanan Manoj[1]*

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India
[2]Present Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

***Correspondence:** Narayanan Manoj (nmanoj@iitm.ac.in)

Mol Phylogenet Evol. 2020 Oct 13;154:106989. https://doi.org/10.1016/j.ympev.2020.106989

Abstract
Histamine receptors belonging to the superfamily of G protein-coupled receptors (GPCRs) mediate the diverse biological effects of biogenic histamine. They are classified into four phylogenetically distinct subtypes H_1 - H_4, each with a different binding affinity for histamine and divergent downstream signaling pathways. Here we present the evolutionary history of the histamine receptors using a phylogenetic approach complemented with comparative genomics analyses of the sequences, gene structures, and synteny of gene neighborhoods. The data indicate the earliest emergence of histamine-mediated GPCR signaling by a H_2 in a prebilaterian ancestor. The analyses support a revised classification of the vertebrate H_3-H_4 receptor subtypes. We demonstrate the presence of the H_4 across vertebrates, contradicting the currently held notion that H_4 is restricted to mammals. These non-mammalian vertebrate H_4 orthologs have been mistaken for H_3. We also identify the presence of a new H_3 subtype (H_3B), distinct from the canonical H_3 (H_3A), and propose that the H_3A, H_3B, and H_4 likely emerged from a H_3 progenitor through the 1R/2R whole genome duplications in an ancestor of the vertebrates. It is apparent that the ability of the H_1, H_2, and $H_{3\text{-}4}$ to bind histamine was acquired convergently. We identified genomic signatures suggesting that the H_1 and H_3-H_4 shared a last common ancestor with the muscarinic receptor in a bilaterian predecessor whereas, the H_2 and the α-adrenoreceptor shared a progenitor in a prebilaterian ancestor. Furthermore, site-specific analysis of the vertebrate subtypes revealed potential residues that may account for the functional divergence between them.

Keywords:
G protein-coupled receptor, histamine, evolutionary history, phylogeny, synteny, convergent evolution

1. Introduction

Histamine, a biogenic neurotransmitter, plays a crucial role in several cellular processes, including inflammatory responses and gastric acid secretion, among others (Hill, 1990; Jones and Kearns, 2011; Kitakaze, 2016; Thurmond et al., 2008). In vertebrates, the diverse physiological effects of histamine are mediated by activation of four metabotropic histamine-specific G protein-coupled receptor (GPCR) subtypes H_1-H_4, each having a different binding affinity for histamine and divergent downstream signaling effects (Arrang et al., 1983; Ash and Schild, 1966; Black et al., 1972; Nakamura et al., 2000). These receptors have distinct expression patterns with both unique and overlapping functions and display constitutive activity as well (Bakker et al., 2000; de Esch et al., 2005; Gantz et al., 1991; Leurs et al., 2005; Liu et al., 2001; Lovenberg et al., 1999; Morisset et al., 2000; Morse et al., 2001; Nguyen et al., 2001; Oda et al., 2000; Smit et al., 1996; Yamashita et al., 1991; Zhu et al., 2001). H_1 and H_2 receptors are expressed in a wide variety of cells, including the

brain and cardiovascular system. The H_3 receptor is found primarily in the central nervous system (CNS) and enterochromaffin cells (Panula et al., 2015). The most recently identified H_4 receptor shares high sequence identity with H_3, but its expression is limited to specific subsets of immune cells (Panula et al., 2015; Zhang et al., 2007). All four subtypes are significant drug targets with highly selective inhibitors for H_1 and H_2, in particular, available for therapy of several disease conditions (Berlin et al., 2011; Leary et al., 2016; Nault et al., 2002; Potnuri et al., 2016; Schneider and Seifert, 2016; Thurmond et al., 2008; Yoshihisa et al., 2017; Zeng et al., 2014; Zhang et al., 2007).

Unlike most aminergic receptor subtypes, the sequence identities between the histamine receptor subtypes are lower, indicating significant divergence. For example, the overall sequence identity between human H_1 and H_2 is about 31 %, while H_1 shares sequence identity in the range 24 – 27% with the H_3 and H_4 receptors. However, H_3 and H_4 receptors share ~42% identity. A search of the human proteome using human the H_1 protein sequence as a query recovered several other aminergic receptors with muscarinic acetylcholine receptors (mAChRs) among the top hits. In contrast, the H_2 receptor hit has a significantly lower E value. Similar results were obtained with the other histamine receptor subtypes. A previous phylogenetic study of H_1, H_2 and H_3 subtypes proposed that these subtypes probably acquired the ability to recognize histamine through convergent evolution (Leurs et al., 2000). It was presumed that invertebrates, in general, lack metabotropic histamine receptors and signal through ionotropic histamine receptors and that metabotropic histamine receptors evolved independently after the divergence of vertebrates from the invertebrates (Roeder, 2003; Steel et al., 1997). However, experimental and gene prediction studies report the presence of putative histamine responsive GPCRs in several invertebrate lineages including snail (*Aplysia californica*) (Carpenter and Gaubatz, 1975), eastern oyster (*Crassostrea virginica*) (Harrison et al., 2015), sea urchin (*Strongylocentrotus purpuratus*) (Leguia and Wessel, 2006; Sutherby et al., 2012), acorn worm (*Saccoglossus kowalevskii*) (Krishnan et al., 2013) and lancelet (*Branchiostoma floridae*) (Burman et al., 2007). One molecular characterization study reports a novel class of histamine receptors unique to the flatworm *Schistosoma mansoni* (El-Shehabi and Ribeiro, 2010; Hamdan et al., 2002). It is interesting to note that protostomal histamine receptors were earlier documented only in Lophotrozhozoa (Carpenter and Gaubatz, 1975; Harrison et al., 2015), although, more recent reports predict their presence in the Ecdysozoa as well (Buckley et al., 2016; Chen and Qian, 2017).

The most recent update on the molecular evolution of the biogenic amine receptors comprised histamine receptors among others but was limited to homologs from the vertebrates (Spielman et al., 2015). The phylogenetic reconstruction carried out by Speilman et al. showed that while the H_3 sequences are broadly distributed across vertebrate taxa, the H_4 clade contained only mammalian sequences. It was proposed that the H_4 subtype is mammalian-specific and arose from a H_3 duplication concurrent with the emergence of the mammals. Given the poor support for the orthologous and paralogous relationships, the presence of multiple paralogous subtypes, and the rapid accumulation of metazoan genomes, the origin and evolution of the histamine receptors require reexamination. Here, we have identified histamine receptor homologs across several metazoan genomes and investigated the molecular evolution of the receptors using a phylogenetic approach combined with gene structure and gene neighborhood synteny analysis. The study provides a revised classification of extant subtypes and a new model of evolution that involves duplications, multiple gene losses and convergence from ancestral amine receptor homologs.

2. Materials and methods

2.1 Identification of histamine receptors

Annotated histamine receptor sequences were first obtained from the GPCR database (GPCRdb) (Munk et al., 2016) and the UniProt protein knowledgebase (UniProtKB). These sequences were used to build and calibrate a hidden Markov model (HMM) profile using HMMBUILD and HMMCALIBRATE programs, respectively, with default parameters. The HMM profile was used as a query (E-value

=0.001) in the HMMSEARCH program to search the 224 proteome datasets downloaded from Ensembl, Joint Genome Institute (JGI) and NCBI genome databases. HMM based programs were run locally using the HMMER suite of programs (Eddy, 1998). An HMM profile built from combining sequences of all four subtypes failed to pick up histamine receptors subtypes as top hits in several cases.

Similarly, BLAST searches using single subtype sequences against the NCBI non-redundant protein database also did not yield histamine subtypes as top hits. For example, using H_1 receptor from human (NP_000852.1) as search seed recovered histamine H_1, mAChR, dopamine, 5-hydroxytryptamine, tyramine and adrenoceptor aminergic receptors, but not all histamine receptor subtypes. In several cases, homologs of other aminergic receptors scored higher than histamine receptors. Hence, separate HMM profiles were built for each subtype and used as search queries and this process yielded putative histamine orthologs as top hits in all searches. A total of 562 sequences from 163 organisms were identified. The total number of entries was refined by removing redundant sequences using CD-hit program at 99% identity (Li and Godzik, 2006). The sequences were also scanned for transmembrane helices using the TMHMM program (Krogh et al., 2001). The sequences that had $\geq$6 and $\leq$7 transmembrane helices were retained while the remaining were treated as fragmentary and removed. The assignment of histamine receptor orthology for putative hits obtained from our searches was considered only if it had known histamine receptors as top hits in a reciprocal BLAST search. The final data set containing a total of 516 sequences from 162 organisms is provided in Table S1. The percentage identity matrix for all paralogs was calculated using SIAS (http://imed.med.ucm.es/Tools/sias.html) server with PID_1 option (Table S2). Isoforms, which are an outcome of alternate splicing events, have been excluded from this study.

2.2 Multiple sequence alignment and phylogenetic tree construction

A protein multiple sequence alignment was generated for the data set that included representatives from histamine and other aminergic receptors using the MAFFT program, with BLOSUM62 as the default scoring matrix and the option E-INS-I was enabled for better accuracy for a data set with multiple conserved domains and long gaps (Yamada et al., 2016). The length of the protein sequences in the multiple sequence alignment ranges from 275 to 705 residues. Most TMs were about 22 residues long. A long third intracellular loop between TM5 and TM6 found in H_1, H_3 and H_4 receptors is relatively short in H_2 receptor subtype (Hill et al., 1997). The N terminal, C terminal and intra-extra-cellular loops (IL/EL) regions show substantial variations, while the seven TM helices are mostly conserved. Phylogenetic trees were constructed using Maximum likelihood (ML) and Bayesian methods. Each tree was run as three sets consisting of (i) full-length alignment (ii) N and C terminal loop regions deleted and (iii) only the 7 TM regions. The TM tree, which had best bootstrap support among the three, was chosen for further analysis. Maximum likelihood approach used to infer phylogeny is as implemented in RAxML 8.2.9 (Stamatakis, 2014) and IQ-TREE (Nguyen et al., 2015; Trifinopoulos et al., 2016). The evolutionary model and parameters for phylogeny implemented in RAxML 8.2.9, containing the sequences in Fig. 1A, B were determined using ProtTest based on the Akaike Information Criterion (minAIC) (Darriba et al., 2011). The best model to determine the evolution of the data set used in this study was determined as Jones-Taylor-Thornton amino acid substitution matrix (Jones et al., 1992) with frequency model along with gamma-distributed (Yang, 1994) for rate among sites (JTT+F+G). Robustness of the tree was tested with 500 bootstrap samplings. Topology of the tree was assessed with bootstrap values mapped on to the initial true tree as implemented in RAxML 8.2.9. The evolutionary model and parameters for sequences in Fig. 1C were determined in IQ-TREE using ModelFinder (Kalyaanamoorthy et al., 2017) based on Bayesian Information Criterion (BIC). The best fit model was determined as JTT+R6, where JTT is the general matrix and R6 represents the FreeRate model with six categories (Soubrier et al., 2012; Yang, 1995). The tree topology was tested using 1000 replicates of ultrafast bootstrap (Minh et al., 2013). FigTree version v1.4.3

software was used to visualize the trees (http://tree.bio.ed.ac.uk/software/figtree/).

Bayesian approach for phylogenetic analysis was performed using MrBayes 3.2.6 (Ronquist and Huelsenbeck, 2003; Ronquist et al., 2012) with gamma-distributed rate variation with frequencies, using JTT model. Posterior probabilities of the trees were approximated using Markov Chain Monte Carlo (MCMC), and analysis was run for 1 x 10^7 generations, and every hundredth tree was sampled. A stop value was applied to terminate the MCMC generations and convergence of MCMC was assessed until the average standard deviation of split frequencies dropped below 0.01 (stop value). The first 25% of sampled trees were disregarded as the *burnin* period so that parameter estimates were only made from data drawn from distributions derived after MCMCs had converged. A consensus tree was built from the remaining 75% of the sampled trees with *sumt* command using 50% majority rule method. The *sump* command was used as a control so that an adequate sample of posterior probability distribution was reached during the MCMC procedure.

2.3 Analysis of gene structure and gene order

The gene structure analysis was carried out using the Gene Structure Display Server 2.0 (GSDS2.0, (Hu et al., 2015). Genomic sequence and the corresponding coding sequence were submitted for generating the exon-intron map with the intron phase information. The results from the server were checked for compatibility with the information present in Ensembl and JGI databases from where the protein sequences are obtained. A total of 384 sequences for which the gene structures compatible with the databases were considered for analysis. Intron positions and phases (GGG phase0, G/GG phase1, and GG/G phase2) were mapped onto the protein sequence alignment. Intron positions were deemed conserved across the alignment for small changes in intron positions (± 10 amino acid residues) (Betts et al., 2001). Distribution of exon-intron length and intron position/phases are given (Data S1 and S2). Synteny analysis was carried out by examining the genes using Genomicus (Louis et al., 2013) for genes retrieved from ensemble database and by manually examining and comparing chromosomal loci using genome browsers for genes obtained from NCBI database.

2.4 Estimation of functional divergence and docking studies

Tests for functional divergence between paralogous subtype groups were performed using the DIVERGE program (Gu et al., 2013). The program was used to identify sites that display Type I and Type II functional divergence. The paralogous groups contain members as annotated in the earlier phylogenetic analysis and are provided in the supplementary file (Tree S1). Statistical strength of Type I and Type II functional divergence was based on the relevant coefficient of divergence (θ), which corresponds to the posterior probability (PP) that a specific site in the alignment had undergone Type I or II divergence (Gu, 1999). These tests for divergence establish whether the θ value is significantly greater than zero, which is an indication of functional divergence. We conducted all possible pairwise comparisons between the major groups to estimate coefficients. The cut-off value of PP was determined by consecutively eliminating the highest scoring sites in the alignment until the coefficient value dropped to zero. Docking studies were carried out using Autodock 4.2.6 (Morris et al., 2009). Histamine was docked with the crystal structure of H_1 histamine receptor (PDB: 3RZE) and the modelled structures of the active form of $H_{2\text{-}4}$ available in GPCRdb (Munk et al., 2016). The four structures were superimposed and the side chains of residues interacting with histamine were oriented with respect to H_1 structure. It should be noted that this docking exercise is with rigid groups and aimed at visualizing the binding of histamine and the differences in the binding sites between subtypes.

2.5 Comparison of substitution rates

Pairwise comparison of substitution rates between histamine receptor subtypes was performed using the codon alignment obtained from RevTrans server. RevTrans server converts a multiple sequence alignment of proteins and the corresponding DNA sequences into a codon alignment (Wernersson and Pedersen, 2003). The phylogenetic tree was constructed for the codon alignment in RAxML 8.2.9 (Stamatakis, 2014)

using a generalized time-reversible substitution matrix along with gamma-distributed and invariant sites (GTR+I+G) as obtained by ProtTest (Darriba et al., 2011). The selection was measured by d*N*/d*S* ratio (ω), where d*N* represents the rate of nonsynonymous substitutions per nonsynonymous site,and d*S* is the rate of synonymous substitutions per synonymous site. When a coding sequence is under negative selection, nonsynonymous substitutions were constrained with respect to the neutral evolution due to their deleterious effect, and therefore $\omega < 1$. Under neutrality, the rate of synonymous substitutions was approximately equal to rate of the nonsynonymous substitutions ($\omega \approx 1$). Alternatively, if the sequence is evolving under positive diversifying selection, d*N* > d*S* and $\omega > 1$. Using the codon alignment and its corresponding tree, we estimated ω using Markov codon models by ML approach as implemented in the CODEML program from PAML 4 software package (Yang, 2007). Optimized log likelihood (lnL) values for both null (fixed ω, lnL0) and alternate model (calculated ω, lnL1)) were calculated. Likelihood ratio tests (LRTs) were performed and significance of the tests was evaluated using the LRT statistic $2*(\text{lnL1}-\text{lnL0}) = 2\Delta L$, which was compared with a chi-square distribution to test whether there is a statistical difference between the null (0) and the alternative (1) models (Yang et al., 2005). We considered a sequence as evolving under positive selection when the LRT for the ω is significant with *P*-value <0.05. The run mode for alternate model (1) was set as model=2 which enables dN/dS ratio calculation, NSsites=2, fix_omega=0, ω=1 and for null model (0) as model=2, NSsites=2, fix_omega=1, ω=1. Codon frequencies were also estimated from the dataset using the F3x4 option.

3. Results and discussion

3.1 Comprehensive survey and overall phylogeny of histamine receptors

3.1.1 A pan-metazoan survey supports the pre-bilaterian origin of histamine receptors

In order to establish the overall phyletic spread and the evolutionary provenance of each histamine receptor subtype, we employed a multi-pronged search strategy in combination with a phylogeny-based approach to assign annotation of the predicted receptors. Recursive searches conducted using discrete HMM profiles for each histamine receptor subtype (see Methods) recovered prototypical versions of these receptors in vertebrates and numerous homologs from a multitude of invertebrate genomes as well as a few homologs from pre-bilaterians. Definitive annotation of these putative histamine receptors and subsequent subtype classification was established by phylogenetic analysis using Maximum Likelihood (ML) and Bayesian approaches. Additionally, the subtype classification was verified using gene structure and gene-neighborhood/synteny analyses.

The phylogenetic tree representing the relationships of the histamine receptors with respect to other aminergic receptors comprised sequences of all histamine receptors and representative receptors comprising α- and β-adrenoceptors, mAChRs, dopamine receptors, and serotonin receptors from a diverse selection of vertebrate and invertebrate genomes (Fig. 1A, Tree S2a). The inclusion of representatives from the other classes of aminergic receptors was based on outcomes of similarity searches using histamine receptors as search seeds against the NCBI non-redundant database. This survey identified a total of 516 histamine receptor sequences from 162 organisms sampled from diverse lineages. Homologs were recovered from representative taxa that contain 47 mammals, 50 birds, 21 fishes, 8 reptiles, 1 amphibian, and 35 invertebrates (Table S1). The comprehensive taxon sampling enabled us to recover several additional putative histamine receptors, including novel homologs from invertebrates. The sample includes homologs from non-bilaterian metazoans like cnidarians and placozoans and, basal bilaterians and invertebrates such as molluscs, annelids, arthropods, echinoderms, hemichordates and cephalochordates. The data implies that the metabotropic histamine receptors are not confined to vertebrates. A previous prediction of histamine receptors in Cnidaria was confirmed here (Anctil, 2009). Nonetheless, our search for orthologs in two of the earliest branching metazoans, namely, sponges and ctenophores, and in some bilaterian lineages such as Nematoda and Tunicata failed to recover

reliable hits that can be deemed canonical histamine receptors.

Most lineages also contain multiple subtype-specific copies (Fig. S1, Tree S2a). These "intra-subtype" paralogs range from two to five copies and vary in terms of their phyletic spread for each subtype and have undergone post-duplication diversification. The sequence-level divergence between these intra-subtype paralogs present across Bilateria suggests that the duplicate copies were probably recruited to carry out additional functional roles post-duplication and have experienced selection (West et al., 1990). In some cases, multiple copies may be attributed to whole-genome duplication (WGD) events, for example, in teleosts. The physiological significance of this copy number specialization remains to be examined.

3.1.2 Histamine receptors are not exclusively lost in the ecdysozoan clade but sporadically distributed

Among the protostomes, histamine receptors were reported to be present only in the annelids, molluscs, Platyhelminthes (lophotrochozoans), and presumed to be lost in arthropods, nematodes and other ecdysozoan lineages (Roeder, 2003; Steel et al., 1997). However, recent reports argue for the presence of histamine receptors in arthropods (Buckley et al., 2016; Chen and Qian, 2017). Recursive searches conducted in our study recovered a total of 14 homologs from diverse panarthropod lineages, and two homologs from a marine protostome *Priapulus caudatus* belonging to Scalidophora clade, which is a sister group to panarthropods (Giribet and Edgecombe, 2017; Misof et al., 2014). Homologs were also recovered from two of the earliest branching arthropods, namely, the crustaceans and chelicerates (Arachnida and Xiphosura) (Misof et al., 2014), and the water-dwelling tardigrades. However, searches spanning hexapods, the largest and the most diverse group of arthropods, recovered only two homologs, specifically from closely related isopterans (termites), suggesting gene loss in the lineage leading to hexapods. The presence of the two termite homologs is either an outcome of cross-species genomic contamination or a recent transfer from other metazoans, although the former scenario appears more reasonable (Koutsovoulos et al., 2016; Merchant et al., 2014). Subsequent recursive searches performed using each of the newly retrieved 16 homologs as search seeds against the entire Ecdysozoa taxa in NCBI NR database did not identify additional homologs. Such sporadic phyletic spread points to two observations: (i) histamine receptors were not exclusively lost in the ecdysozoan clade after their split from the common ancestor of the Lophotrochozoa and Ecdysozoa; (ii) multiple lineage-specific loss of the receptor occurred within diverse ecdysozoan lineages and the receptor is entirely lost in the lineage leading to the nematodes.

3.1.3 Histaminergic receptors in urochordates

The presence of canonical histamine receptors in the tunicates, a member of the Urochordata, is still debated. A recent study reported the characterization of an H_2 receptor in the tunicate *Botryllus schlosseri* based on its clustering with a predicted H_2 receptor sequence from *Ciona intestinalis* (XP_002125985.1) (Cima and Franchi, 2016). However, other exhaustive comparative studies of the genome and the GPCR repertoire of *C. intestinalis* report the absence of histamine receptors in the tunicate lineage (Burke et al., 2006; Dehal et al., 2002; Kamesh et al., 2008). Therefore, we examined the two tunicate sequences in more detail. Neither of these sequences recovers histamine receptors as their top hits in BLASTP searches against the NCBI NR database, and the annotation could not be validated using phylogenetic analysis. We cannot rule out the possibility that the predicted *B. schlosseri* receptor may be a novel amine receptor that can be stimulated by histamine. Given the presence of histidine decarboxylase and its reaction product, histamine, in both nematodes and tunicates, it is tempting to speculate that the histamine signaling is either carried by a closely related aminergic receptor or a more derived version of the canonical type histamine receptor, similar to the new class of histamine receptor reported in the flatworm (El-Shehabi and Ribeiro, 2010; Hamdan et al., 2002). In order to detect different versions of histamine receptors, we used the flatworm and arthropod sequences as search seeds in NCBI NR search and examined the orthology of the top hits in the phylogenetic analysis. However, the clustering patterns were ambiguous. It is likely that some lineages may

indeed contain novel classes of metabotropic aminergic receptors that may be stimulated by histamine and remain to be experimentally characterized. Alternatively, histamine signaling in lineages that lack metabotropic histamine receptors may be ionotropically mediated (Akashi et al., 2018; Cebada and Garcia, 2007; Gisselmann et al., 2002; Kita et al., 2017).

3.1.4 Overall phylogenetic relationship of histamine and other aminergic receptors

It should be noted that histamine receptors display higher "inter-family" sequence identities (with other aminergic receptors) compared to that for "intra-subtype". This pattern and the fact that the time of divergence of several aminergic receptor families and histamine receptor subtypes likely predate the emergence of Bilateria presents a considerable challenge to obtain well-supported subtype-specific monophyletic clades in phylogenetic reconstruction. Hence, we reconstructed phylogenetic trees using different alignment options that include complete sequences or partial sequences containing the transmembrane (TM) regions alone or TMs containing intra- and extra-cellular loops (IL/EL) and lacking either termini (see methods). The tree with most well-resolved clades was selected for further analysis. The alignment that contained the TM and the IL-EL regions and lacked both the termini yielded trees with the best confidence values for the nodes within. Both ML and Bayesian phylogenetic tree construction approaches yielded almost identical topologies for the well-supported clades. The Bayesian tree was chosen for further analysis. A separate phylogenetic tree was generated for the histamine receptors alone for comparison.

Overall, the aminergic receptors and their subtypes cluster into distinct clades that do not correspond fully to the ligand-binding property (Fig. 1A and B, Tree S2a and S2b). For instance, the histamine and adrenoceptor subtypes form paraphyletic groups. Strongly supported clades with Bayesian posterior probability (BPP) >96% separate the histamine receptor subtypes, with the invertebrate orthologs placed basal to their vertebrate counterparts, largely following the general archetype of the tree of life. As expected, the H_3 and H_4 subtypes form a well-supported monophyletic H_3-H_4 clade (Coge et al., 2001a; Coge et al., 2001b; Liu et al., 2001; Oda et al., 2000; Tardivel-Lacombe et al., 2001). Furthermore, the H_3-H_4 subclade is a sister clade to the mAChR clade and the monophyletic H_1 subtype clade appears as a sister clade to that containing H_3-H_4 and mAChRs. The tree topology is broadly congruent with that in an earlier study carried out using sequences from vertebrates (Spielman et al., 2015). Interestingly, flatworm histamine receptors (Smp_043260, Smp_043290, Smp_043340), previously characterized as a new class of histamine receptors, are placed basal to H_1, H_3-H_4, and mAChR receptors clade with 99% BPP (El-Shehabi and Ribeiro, 2010; Hamdan et al., 2002; Zamanian et al., 2011). This pattern suggests that these are highly diversified versions of ancestral histamine receptors already present at the base of the Bilateria, although they may also have independently evolved the ability to bind histamine. The clustering pattern also suggests that the predicted flatworm histamine receptor (Smp_126730) is perhaps a serotonin receptor.

A striking observation is that the H_2 formed a distinct clade separate from the other subtypes. Serotonin and α-adrenoceptors are placed basal to the clade constituting H_3, H_4, mAChRs, and H_1, forming one major clade (98% BPP), whereas the H_2 clade is clearly an outgroup here. The data thus support the hypothesis that the H_1, H_2 and H_3-H_4 subtypes may have been derived independently from closely related aminergic receptors present in an ancestral bilaterian and subsequently acquired convergent changes that enabled the binding of histamine. In general, within our comprehensive dataset, subtypes H_1, H_2 and H_3-H_4 share higher sequence identity with other aminergic GPCRs, than between them (Table S3), in agreement with an earlier evolutionary analysis with a limited dataset (Leurs et al., 2000). Indeed, the first homology-based attempt to clone the H_3 based on H_1 and H_2 subtype sequences were unsuccessful due to the poor sequence relationship (Lovenberg et al., 1999).

The high sequence similarity between the H_3 and H_4 strongly suggests a close evolutionary relationship. The presently accepted notion that the H_4 emerged in the mammalian lineage by duplication from an H_3 progenitor is supported by the phylogenetic reconstruction of vertebrate

amine receptors (Spielman et al., 2015). The overall topology pattern of the histamine receptor subtypes obtained from our comprehensive dataset is in line with the previous study (Spielman et al., 2015) (Fig. 1A). However, the topology of the H_3 and H_4 clades here is inconsistent with the duplication occurring in an ancestor of the mammalian lineage (Fig. 1B, Tree S2a). First, there appear to be multiple duplication events within the H_3. The deuterostomal invertebrate group, including the echinoderms, hemichordates and the cephalochordates, contain up to five copies of the H_3 that seem to be local duplications within each species. Second, in most vertebrates, including the mammals, there are two or three copies of the H_3 subtype. Moreover, these vertebrate H_3 homologs are paraphyletic, with well-supported clades (clades I, II, III, in Fig. 1B), each containing H_3 homologs across the vertebrates and each consistent with the species tree. The data suggest duplication and specific diversification within the vertebrate H_3 subtype. Most notably, the placement of mammalian H_4 clade within the vertebrate H_3 clade III, and not basal to the mammalian H_3 clade, is unexpected and surprising. For instance, several predicted H_3 homologs from non-mammalian vertebrates clustered basal to the mammalian H_4 clade forming a larger clade with high confidence support (89% BPP). The clustering pattern in Fig. 1B leads to an interpretation of an intriguing possibility that the non-mammalian vertebrate H_3 sequences and the mammalian H_4 sequences in clade III shared a last common ancestor and are divergent from the paralogous vertebrate H_3 sequences in the clades I and II. Although this inconsistency can be attributed to the limitations of phylogenetic analysis when dealing with rapidly diverging duplicate sequences, these observations also point to some unsettled queries: (i) are H_4 receptors exclusive to eutherian mammals as previously thought or do they have a deeper origin in vertebrates? (ii) if so, can the duplication event that gave rise to the H_4 receptor be dated? (iii) do vertebrates indeed harbor distinct paralogous H_3 subtypes? In order to further investigate the H_3-H_4 relationships, a new phylogenetic tree was reconstructed solely using these H_3 and H_4 homologs (Fig. 1C). The topology of the major clades here is better supported and is mostly congruent with the tree containing multiple aminergic receptor families. It is to be noted that the subtype annotation of receptors in Fig. 1A and B, follows the original annotation whereas Fig. 1C follows a revised nomenclature that resulted from an analysis of gene neighborhoods as described in the section below.

3.2 Synteny analysis supports presence of distinct H_3 paralogs and emergence of H_4 in the vertebrates

To obtain an independent line of evidence for evolutionary relationships among the subtypes, we examined gene neighborhoods for the regions harboring the four genes (Data S3). Conserved gene neighborhood between distantly related organisms is indicative of selection pressures. It is often used to substantiate phylogenetic inferences as well as to explain inconsistencies in gene family clustering that may arise from the limitations of methodologies employed for tree reconstruction. Indeed, synteny analyses have been used earlier in reconstructing the evolutionary histories of several gene families, including GPCRs (Gao et al., 2018; Ocampo Daza and Larhammar, 2018; Pedersen et al., 2018, 2019).

The four histamine receptor subtype genes have different chromosomal loci. For example, human genes *HRH1*, *HRH2*, *HRH3*, and *HRH4* are located on chromosomes 3, 5, 20, and 18, respectively. Within each of the H_1, H_2, and H_3-H_4 subtypes, the conservation of gene order was high in mammalian, bird, reptile, amphibian, and fish lineages (Data S3). However, the neighborhood is not conserved across the subtypes except between H_3 and H_4, which share several neighboring genes. All mammalian H_4 homologs are bordered by paralogous genes such as *IMPACT*, *OSBPL1A* and *LAMA3* on the 5′ end, and *ZNF521*, *SS18*, *PSMA8*, *TAF4B, CDH2* on the 3′ end. Paralogous members of these gene families, including *OSBPL*, *LAMA*, *SS18L*, *PSMA*, *TAF* and *CDH* are also flanking genes for two among the three clades of duplicate vertebrate H_3 sequences (clades I and III, Fig. 1B, Fig. 2). This strongly suggests that the conserved synteny between these H_3 and H_4 genomic regions is a result of a shared duplication. In particular, the *IMPACT* gene present tandem to

HRH4 is a subtle yet unambiguous synapomorphy to distinguish the vertebrate H_4 paralogon from the vertebrate H_3 paralogon.

The phylogenetic reconstruction together with the synteny data clearly indicate that the non-mammalian vertebrate H_3 sequences in clade III of Fig. 1B that clustered basal to the mammalian H_4 are true H_4 receptors, and not H_3 as misannotated currently in sequence databases. The synteny data also indicates that the H_3 Fishes and H_3 Amphibians in Fig. 1B, that are sister clades to clade III, are also true H_4 and not H_3. Moreover, clades I and II represent two distinct and hitherto unrecognized H_3 paralogs within the vertebrates (Fig. 2). Therefore, we propose a revised nomenclature for the H_3-H_4 receptors in the vertebrates (Fig. 1C). First, the non-mammalian H_4 vertebrate homologs that have been mistaken for H_3 are now reannotated as H_4 (Fig. 1C, clade III). Next, the canonical vertebrate H_3 paralogs of clade I are named H_3A. The gene neighborhood of the vertebrate H_3 paralogs in clade II is not syntenic with H_3A/H_4 genomic region and thus represents a novel H_3 receptor subtype, which we name H_3B (Fig. 2). Lastly, the invertebrate H_3 does not share gene neighborhood synteny either with the H_3A/H_4 or H_3B subtypes and as expected, is placed basal to the vertebrate H_3-H_4 clade and retains the name H_3. In order to benefit researchers with their studies based on the revised classification, we provide a general scheme of nomenclature where local duplicate copies within subtypes include a subscript lowercase alphabet, for example, H_1 H_{1b}, H_{1c} (Table S1). The full list of the invertebrate H_3 sequences, the vertebrate H_3A and H_3B paralogs, and the expanded list of vertebrate H_4 orthologs indicating their current and revised classification are provided in Table S4. The overall sequence identities within and between the subtypes is provided in Table S5.

The fact that the H_4 gene emerged much earlier than was hitherto recognized also allows us to date the H_3-H_4 split. This split most likely occurred in an ancestor of the vertebrates since the H_4 was identified in the lamprey (Cyclostomata), a jawless vertebrate. Although this sequence exhibited discrepancies in its placement in different trees (for example, it clustered basal to H_3B, Fig. 1C), the H_4 subtype classification is validated using the *IMPACT* gene synapomorphy. The lamprey presents a unique situation lacking both H_3A and H_3B subtypes. A search in lamprey for genes present in the jawed vertebrate H_3A paralogon did not pick up a contiguous region that could be examined for synteny, although it yielded reliable hits for at least two genes (*SS18L1* and *CABLES1*) present in the H_3A paralogon. Therefore, H_3A either is truly lost in lamprey or is a consequence of sequence assembly anomalies arising out of the unique repetitive nature of the genome (Smith et al., 2009) (Qiu et al., 2011; Smith et al., 2013) (Data S3, Fig. 2). It is noteworthy that the H_3B paralogon in lamprey is indeed conserved across several vertebrates, although the H_3B gene itself is lost. The presence of this paralogon in lamprey, therefore, suggests a deeper origin of the H_3B subtype, likely predating the lamprey-gnathostome divergence (Data S3, Fig. 2). Additionally, unlike the H_3A subtype that is largely ubiquitous across vertebrates, the H_3B displays sporadic phyletic distribution. Orthologs are present in some ray-finned fish, and, in reptiles, birds and amphibians. However, in mammals, the subtype is absent in the monotremes and present in the marsupials and, among the Eutheria, the H_3B subtype is present solely in the Glires and lost in all other lineages (Data S3, Fig. 2). Interestingly, the Actinopterygiian H_3A paralog is present only in Holostei and absent in Teleostei. The cartilaginous fish (Chondrichthyes) is the sole vertebrate lineage that lost the H_4 subtype, although the gene neighborhood is conserved (Data S3, Fig. 2). Finally, the intron/exon patterns of the genes were also examined for conservation patterns. In general, gene structures were mostly conserved within each subtype, but not across subtypes, except between the vertebrate H_3A, H_3B and H_4, which share conserved intron positions and phases, as expected (Fig. S2).

Our demonstration of the revised origin and divergence of the H_3 and H_4 has significant implications for functional studies in vertebrate lineages. It is pertinent to examine why the origin, divergence and classification of the H_3/H_4 were previously mistaken. It is most likely that the one-to-one orthologous and paralogous relationships between the H_3 duplicates and the H_4 could not be captured because of the high pairwise similarities between them and as a result these homologs

have tended to be named as H_3-like or H_4-like in sequence databases. Furthermore, previous phylogenetic studies of histamine receptor datasets have overlooked the well-supported cluster of H_3-H_4 sequences and by design, did not explore the evolutionary relationship between these two subtypes. In particular, the synteny analysis provided the most definite line of evidence that allowed us to propose the revised classification.

3.3 Shared ancestry of H_1, H_{3-4} and mAChRs and independent origin of H_2

Next, we sought to examine conserved gene neighborhood patterns, if any, that reflect the phylogenetic clustering of H_1, H_3-H_4 with mAChRs. The mAChRs have poor shared synteny across all five subtypes, although there is strong evidence for divergent evolution from a single ancestral mAChR gene (Pedersen et al., 2018). The H_1, H_3-H_4 subtypes share some neighboring genes with the mAChRs (Data S3). For instance, genes for the *SLC* (Solute carrier family protein) are shared between H_1 and mAChR1, 2, 5, while the *ATG* (Autophagy related protein) is shared between H_1 and mAChR4. H_4 and mAChR4 have *ZNF* (zinc finger protein) in common, while *TAF* (TATA-box binding protein associated factor) is common between H_3A, H_4 and mAChR1. Interestingly, in the echinoderm starfish, the mAChR2 and dopamine receptors occur in tandem with two copies of H_3 (Fig. S3). One interpretation is that the mAChR and the ancestral H_3 genes were generated out of tandem duplication of a progenitor gene in an ancestor of the Bilateria, as indicated by the node at which these branches separate. Based on the findings so far, it is tempting to speculate that the phylogenetically related mAChR, H_1, and the H_3-H_4 genes originated from a single bilaterian progenitor gene by local duplication and evolved by accumulated mutations independently (Fig. 1A, Fig. 3). Next, our phylogenetic analyses and gene presence data carried out on an expanded genome dataset demonstrates an earlier origin of the H_2 in a prebilaterian ancestor (sea anemone, hydra, Staghorn coral and trichoplax).

The exact order of the emergence of these paralogous genes arising out of tandem duplication events cannot be estimated reliably using the current data. However, patterns of tandem prevalence, shared synteny and phylogenetic relationships lend robust support to the previously known hypothesis that the H_2 evolved histamine binding independently of the H_1 and H_{3-4}. While there is always the possibility that the tandem occurrences identified above are convergent phenomena in the specific genomes, it is well known that the expansion of several families/subfamilies in the GPCR superfamily have emerged through tandem duplications during metazoan evolution prior to the 1R/2R WGD (Hwang et al., 2013; Kim et al., 2014; Larhammar and Salaneck, 2004; Sefideh et al., 2014; Tostivint et al., 2014; Yun et al., 2015). An extensive comparative analysis of these phylogenetically related aminergic receptors at the levels of sequence and gene organization in a large dataset that includes diverse invertebrate genomes may resolve the evolutionary hierarchy and the likely path of the provenance of histaminergic signaling. For instance, if mAChRs were found to be derived from H_1 early in metazoan evolution, then such "family/ligand" switching might have happened twice during metazoan evolution since H_{3-4} were probably derived from mAChRs. Similar convergent evolution in the ability of subtypes to bind the same agonist has earlier been established in the dopamine receptor and serotonin receptor (Callier et al., 2003; Yamamoto and Vernier, 2011). Finally, it is remarkable that non-classical histamine recognition also evolved independently in a clade of olfactory trace amine-associated receptors (TAARs) by the acquisition of appropriate amine recognition by a salt bridge interaction with an aspartic acid residue, illustrating how divergent GPCRs can evolve pockets for the same agonist (Li et al., 2015).

3.4 Structural comparison of histamine and muscarinic receptor subtypes

Our evolutionary model suggests that the H_3-H_4 and H_1 and the mAChRs most likely shared a most recent common ancestor. Since crystal structures of human H_1 (PDB: 3RZE) and muscarinic M_{1-4} (PDB: 5CXV, 3UON, 4U15, 5DSG) are available, we compared the active sites of these for shared and divergent features that may shed light on the evolution of different ligand specificities (supplementary note).

Remarkably, a significant number of residues that interact with the ligands were completely conserved between the histamine receptors and mAChRs (Fig. S4a and S4b, Table S6). One striking difference in the active site residues between histamine receptors and mAChRs is the presence of a conserved Asn residue in mAChRs ($N382^{6.52}$:M_1) (Ballesteros-Weinstein number in superscript (Ballesteros and Weinstein, 1995)) which makes a hydrogen bond with the ligand. The structurally equivalent residue is Phe in H_1 ($F432^{6.52}$:H_1) which makes hydrophobic interactions with the ligand and is not conserved in histamine receptors (Fig S4a).

3.5 Evolutionary history of histamine receptors

Given the large repertoires of the GPCR superfamily genes in metazoans, it is clear that phylogenetic approaches alone will be insufficient to unravel the complex evolutionary history of families and subfamilies within. Our study forces a significant revision of the existing model for the origin and evolution of the histamine receptors. The tracing of a prebilaterian origin of GPCR mediated histamine signaling, the conclusion that the H_4 being present across vertebrates, and the recognition of the new H_3B are fundamental discoveries here. Our data provide an updated phyletic distribution, a revised classification, and an overall evolutionary history of the receptors to establish a more reliable framework for the design of experiments and interpretation of results from comparative studies of function (Fig. 3).

The earliest emergence of a canonical histamine receptor ortholog in a pre-bilaterian ancestor of the Placozoa, Cnidaria and Bilateria appears to be that of the H_2 (~620 mya). Orthologs of H_1 and H_3 in the protostome and deuterostome lineages indicate that these probably first emerged independently of H_2 in the Bilateria, subsequent to its divergence from the ParaHoxozoan ancestor. The H_3 duplicates appear to have first occurred in the deuterostomal lineage and are largely restricted to the invertebrates in this lineage. The phyletic spread of the receptors across multiple bilaterian lineages includes possible lineage-specific losses in hexapods, nematodes and tunicates. The H_4 and H_3A subtype and the distinct H_3B subtypes emerged as early as the jawless vertebrates (Fig. 3). The origin and evolution of the extant H_3 and H_4 subtypes appear to follow the classical expansion following the two rounds (1R and 2R) of whole-genome duplication in an early vertebrate ancestor. In this scheme, an ancestral paralogon that harbored the H_3, probably duplicated under 1R to give rise to H_3B-like and H_3A/H_4-like progenitors. The 2R expansion generated the H_3A and H_4 from the H_3A/H_4-like progenitor and the H_3B from the H_3B-like progenitor, followed by a secondary loss of the H_3B duplicate (Fig. 3). The loss of synteny between the H_3B paralogon and the H_3A/H_4 paralogon could be an outcome of chromosomal translocation/rearrangements or a single gene translocation that occurred after the WGD. Alternatively, under a less parsimonious model, each of the three vertebrate H_3B, H_3A, and H_4 may perhaps have evolved independently from separate duplicate copies of H_3 already present in a deuterostomal ancestor (Fig. 3). This scenario requires persistent losses of vertebrate paralogs that were generated from either local gene duplications or the 1R/2R WGD events.

3.6 Implications for H_3A, H_3B, H_4 functional studies

The relative functional status of these duplicated genes can be elucidated from their presence/absence in vertebrates. Presence of H_3A and H_4 are evidently under strong negative selection since they are present in all vertebrate lineages. Presumably, the H_3A and H_4 have undergone neofunctionalization (NF), where one copy may acquire a novel function while preserving the ancestral function in the other copy. In contrast, the sporadic loss of H_3B across vertebrates suggests that this subtype might have undergone subfunctionalization (SF) with respect to its ancestral H_3 gene and hence may be expendable in several lineages. Since many species have retained the H_3B paralogs, it is unlikely that the H_3B is redundant (Fig. 3) and probably reflects the fixation of new adaptive functions that remain to be established in these lineages. Most reported functional/biochemical characterization in literature corresponds to that of the H_3A paralog in mammalian models like human, mice and rat. Although the unique H_3B paralog has not been recognized earlier, there is

limited functional data from studies of this paralog in zebrafish, a teleost that lacks the canonical H_3A. For instance, *hrh3B* knock-out zebrafish larvae display essentially normal sleep/wake cycle in contrast to the *hrh3A* mutant mouse, which exhibits sleep fragmentation at night (Chen et al., 2017; Gondard et al., 2013). In addition, *hrh3A* tends to be expressed in mammalian hindbrain regions, whereas the *hrh3B* paralog is not expressed in the zebrafish (Puttonen et al., 2018). Therefore, these past works imply some level of functional specialization between the two H_3 subtypes. Careful design of functional experiments on lineage-specific knock-outs, for instance, on those that contain either or both H_3 subtypes, are needed to characterize and dissect distinct functional effects, if any, on the respective phenotypes. Notably, the spotted gar (Holestei), the earliest ray-finned fish that did not undergo the teleost genome duplication (TGD), possesses the full complement of the H_3A, H_3B and H4 ohnologs that are lost reciprocally in several teleosts and tetrapods. We propose that the gar, a "living fossil", appears to be an ideal model system to elucidate the mechanisms and effects of functional partitioning of ancient versions of these subtypes and guide further investigations into the evolution of histaminergic systems in tetrapods, including humans.

3.7 Functional divergence and substitution rates between histamine receptor subtypes

3.7.1 Estimation of functional divergence and docking studies

Next, we examined the effect of selection and searched for evolutionary determinants of functional divergence between sets of subtypes. Tests were carried out to obtain estimates for substitution rates and site-specific rate shifts (supplementary note) using the revised subtype annotation in our expanded dataset of vertebrate sequences. The amino acid sites responsible for functional divergence (Type I and Type II) were predicted. Between pairs of paralogous groups, amino acid sites that are highly conserved in one and are variable in the other indicate Type I divergence. An amino acid site that was highly conserved within groups but had radically different properties between the groups indicates Type II divergence. The divergence value (θ), Z score and *P*-value for all comparisons are given in Table S7. Pairwise comparisons show that the θ values are significantly much higher than zero (*P*-value < 0.02) for several comparisons. For instance, the NF between the H_3 and H_4 pair because of changes in selective constraint at individual sites is evident from a significant Type I coefficient value (θ) of 0.47. Posterior probability (PP) analyses predicted several significant Type I and Type II sites (PP values, 0.90 - 0.99) that may contribute to functional divergence between the subtypes.

A total of 86 Type I and 65 Type II divergent sites were identified from all pairwise comparisons (Tables S8 and S9). Interestingly, several predicted sites have pharmacological evidence for their functional roles. For example, Type I residue $K191^{5.40}$ of human H_1 affects agonist and antagonist binding and confers H_1 selectivity (Leurs et al., 1995; Leurs et al., 1994; ter Laak et al., 1995; Wieland et al., 1999). This site is also responsible for the formation of partial bidentate hydrogen bonding with imidazolyl and similar group containing ligands (Seifert et al., 2013). This site is conserved in H_1 and divergent in H_4 (Table S8). Among Type II sites, residues $N198^{5.461}$ of H_1 (Leurs et al., 1994; Moguilevsky et al., 1995; Ohta et al., 1994) and $E206^{5.461}$ in human H_3 (Uveges et al., 2002) and $E182^{5.461}$ in human H_4 (Kiss et al., 2008; Shin et al., 2002) are known to affect histamine binding and determine subtype selectivity (Seifert et al., 2013). The list of predicted sites for which putative functional roles are supported by biochemical evidence is given (Table S10). The sites were demarcated by their location in the TM, IL and EL regions and plotted on a snake plot (Fig. S5, Tables S8 and S9) to understand their roles in ligand binding and signal transduction. The sites were also mapped onto structural models of the subtypes docked with histamine. As expected, several residues are positioned in the cavity that constitute the putative histamine binding sites and therefore provide a structural rationale for the divergent modes of histamine binding and activation between the subtypes (supplementary note, Fig. S6).

3.7.2 Comparison of substitution rates

Positive selection was detected between two comparisons, H_1vsH_2 and H_3vsH_4 (ω = 3.53, ω =

31.75) with significant values for LRT (Table S11). The Bayes Empirical Bayes (BEB) method was used to identify 13 residues under positive selection (Table S12 and supplementary note). The sites were marked on the structure of human H_1 (PDB: 3RZE) according to their corresponding positions in the multiple sequence alignment (Fig. S9). For instance, three sites were found in the extracellular side of the TM of which $Y108^{3.33}$ of human H_1 interacts with the ligand doxepin without disrupting the phosphate ion which is present in the ligand-binding pocket. This anion binding site in the ligand-binding pocket is unique to this H_1 doxepin complex and is supported by the fact that phosphate ion affects the stability of the H_1 receptor and the binding of ligands (Shimamura et al., 2011). $Y108^{3.33}$ constitutes the lipophilic binding cavity, and its role in binding of H_1 antagonists is supported by mutational data (Kiss et al., 2004). Furthermore, it is interesting to note that only these two predicted sites, human H_3 $K108^{3.26}$, and human H_1 $L207^{5.55}$, were common between functional divergence and positive selection analyses. We conclude that the functional variations within the histamine receptors may be primarily dictated by specific sites or a combination of sites predicted in this study. The results may be used for the rational design of agonist binding studies, site-directed mutagenesis and other experiments aimed at developing novel and effective drugs targeting different histamine receptors.

4. Conclusions

Our bioinformatics analyses suggest that the four histamine receptors H_1-H_4 show wide phyletic spread in metazoans, and its emergence can be traced back to the last common ancestor of parahoxozoans, much earlier than previously reported. The previously recognized canonical H_3 subtype is, in fact, two distinct subtypes, H_3A and H_3B. Further, the origin of the H_4 in a vertebrate ancestor dispels the currently held notion that they are restricted to mammals. In particular, the H_4 is the most recent histamine receptor to be identified, with significant questions with respect to its functional and pharmacological properties. This study also provides a revised evolutionary model for the origin and divergence of the H_3-H_4 subtypes in the vertebrates arising via WGDs. We detected genomic signatures of convergent evolution of the subtypes in acquiring histamine binding ability. It is likely that the H_1 and H_3-H_4 shared a recent common ancestor with the mAChRs, while the H_2 shared a recent common ancestor with the α-adrenoceptor. Site-specific analysis demonstrates divergence between the histamine receptors and indicates that a majority of functionally divergent sites are located at TM3, 4, 5, 6 and extracellular loops which affect binding of specific agonists and antagonists. Presumably, minor differences in the active site cleft of these receptors can lead to pronounced functional divergence and the subsequent emergence of a distinct family or subfamily. These findings will advance understanding of histamine receptor function with regard to pharmacological and physiological relevance of specificity, activity, selectivity and cross-reactivity of the numerous known agonists and antagonists.

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

Acknowledgements

We acknowledge infrastructural and financial support from the Bioinformatics Infrastructure Facility, supported by the Department of Biotechnology (DBT), Government of India (GoI). We acknowledge infrastructural support from the High-Performance Computing Environment, IIT Madras. ISR thanks Department of Science and Technology (DST-INSPIRE), India, for providing research fellowship. This work was supported by the DBT, GoI, [Grant BT/PR11748/BID/07/285/2008].

References

Akashi, H.D., et al., 2018. Physiological responses of ionotropic histamine receptors, PxHCLA and PxHCLB, to neurotransmitter candidates in a butterfly, *Papilio xuthus*. J Exp Biol.

Anctil, M., 2009. Chemical transmission in the sea anemone *Nematostella vectensis*: A genomic perspective. Comp Biochem Physiol Part D Genomics Proteomics 4, 268-289.

Arrang, J.M., et al., 1983. Auto-inhibition of brain histamine release mediated by a novel class (H3) of histamine receptor. Nature 302, 832-837.

Ash, A.S., Schild, H.O., 1966. Receptors mediating some actions of histamine. Br J Pharmacol Chemother 27, 427-439.
Bakker, R.A., et al., 2000. Constitutive activity of the histamine H(1) receptor reveals inverse agonism of histamine H(1) receptor antagonists. Eur J Pharmacol 387, R5-7.
Ballesteros, J.A., Weinstein, H., 1995. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods in neurosciences. Elsevier, pp. 366-428.
Berlin, M., et al., 2011. Histamine H3 receptor as a drug discovery target. J Med Chem 54, 26-53.
Betts, M.J., et al., 2001. Exon structure conservation despite low sequence similarity: a relic of dramatic events in evolution? EMBO J 20, 5354-5360.
Black, J.W., et al., 1972. Definition and antagonism of histamine H2-receptors. Nature 236, 385-390.
Buckley, S.J., et al., 2016. In silico prediction of the G-protein coupled receptors expressed during the metamorphic molt of *Sagmariasus verreauxi* (Crustacea: Decapoda) by mining transcriptomic data: RNA-seq to repertoire. Gen Comp Endocrinol 228, 111-127.
Burke, R.D., et al., 2006. A genomic view of the sea urchin nervous system. Dev Biol 300, 434-460.
Burman, C., et al., 2007. Eleven new putative aminergic G-protein coupled receptors from Amphioxus (*Branchiostoma floridae*): identification, sequence analysis and phylogenetic relationship. Invert Neurosci 7, 87-98.
Callier, S., et al., 2003. Evolution and cell biology of dopamine receptors in vertebrates. Biol Cell 95, 489-502.
Carpenter, D.O., Gaubatz, G.L., 1975. H1 and H2 histamine receptors on Aplysia neurones. Nature 254, 343-344.
Cebada, J., Garcia, U., 2007. Histamine operates Cl -gated channels in crayfish neurosecretory cells. J Exp Biol 210, 3962-3969.
Chen, A., et al., 2017. Genetic Analysis of Histamine Signaling in Larval Zebrafish Sleep. eNeuro 4.
Chen, L., Qian, P.Y., 2017. Review on Molecular Mechanisms of Antifouling Compounds: An Update since 2012. Mar Drugs 15.
Cima, F., Franchi, N., 2016. Histamine Stimulates Ciliary Beat Frequency via the H2 Receptor in the Protochordate *Botryllus schlosseri*. J Exp Zool B Mol Dev Evol 326, 176-192.
Coge, F., et al., 2001a. Genomic organization and characterization of splice variants of the human histamine H3 receptor. Biochem J 355, 279-288.
Coge, F., et al., 2001b. Structure and expression of the human histamine H4-receptor gene. Biochem Biophys Res Commun 284, 301-309.
Darriba, D., et al., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164-1165.
de Esch, I.J., et al., 2005. The histamine H4 receptor as a new therapeutic target for inflammation. Trends Pharmacol Sci 26, 462-469.
Dehal, P., et al., 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298, 2157-2167.
Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755-763.
El-Shehabi, F., Ribeiro, P., 2010. Histamine signalling in *Schistosoma mansoni*: immunolocalisation and characterisation of a new histamine-responsive receptor (SmGPR-2). Int J Parasitol 40, 1395-1406.
Gantz, I., et al., 1991. Molecular cloning of a gene encoding the histamine H2 receptor. Proc Natl Acad Sci U S A 88, 429-433.
Gao, T., et al., 2018. The Neuropeptide Y System Regulates Both Mechanical and Histaminergic Itch. J Invest Dermatol 138, 2405-2411.
Giribet, G., Edgecombe, G.D., 2017. Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. Integr Comp Biol 57, 455-466.
Gisselmann, G., et al., 2002. Two cDNAs coding for histamine-gated ion channels in *D. melanogaster*. Nat Neurosci 5, 11-12.
Gondard, E., et al., 2013. Enhanced histaminergic neurotransmission and sleep-wake alterations, a study in histamine H3-receptor knock-out mice. Neuropsychopharmacology 38, 1015-1031.
Gu, X., 1999. Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16, 1664-1674.

Gu, X., et al., 2013. An update of DIVERGE software for functional divergence analysis of protein family. Mol Biol Evol 30, 1713-1719.
Hamdan, F.F., et al., 2002. A novel *Schistosoma mansoni* G protein-coupled receptor is responsive to histamine. Mol Biochem Parasitol 119, 75-86.
Harrison, J., et al., 2015. The Presence of Histamine and a Histamine Receptor in the Bivalve Mollusc, *Crassostrea virginica*. In Vivo 36, 123-130.
Hill, S.J., 1990. Distribution, properties, and functional characteristics of three classes of histamine receptor. Pharmacol Rev 42, 45-83.
Hill, S.J., et al., 1997. International Union of Pharmacology. XIII. Classification of histamine receptors. Pharmacol Rev 49, 253-278.
Hu, B., et al., 2015. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31, 1296-1297.
Hwang, J.I., et al., 2013. Expansion of secretin-like G protein-coupled receptors and their peptide ligands via local duplications before and after two rounds of whole-genome duplication. Mol Biol Evol 30, 1119-1130.
Jones, B.L., Kearns, G.L., 2011. Histamine: new thoughts about a familiar mediator. Clin Pharmacol Ther 89, 189-197.
Jones, D.T., et al., 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8, 275-282.
Kalyaanamoorthy, S., et al., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587-589.
Kamesh, N., et al., 2008. The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. BMC Evol Biol 8, 129.
Kim, D.K., et al., 2014. Coevolution of the spexin/galanin/kisspeptin family: Spexin activates galanin receptor type II and III. Endocrinology 155, 1864-1873.
Kiss, R., et al., 2004. Homology modelling and binding site mapping of the human histamine H1 receptor. Eur J Med Chem 39, 959-967.
Kiss, R., et al., 2008. Binding mode analysis and enrichment studies on homology models of the human histamine H4 receptor. Eur J Med Chem 43, 1059-1070.
Kita, T., et al., 2017. Pharmacological characterization of histamine-gated chloride channels from the housefly *Musca domestica*. Neurotoxicology 60, 245-253.
Kitakaze, M., 2016. Clinical Evidence of the Role of Histamine in Heart Failure. J Am Coll Cardiol 67, 1553-1555.
Koutsovoulos, G., et al., 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proc Natl Acad Sci U S A 113, 5053-5058.
Krishnan, A., et al., 2013. Remarkable similarities between the hemichordate (*Saccoglossus kowalevskii*) and vertebrate GPCR repertoire. Gene 526, 122-133.
Krogh, A., et al., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305, 567-580.
Larhammar, D., Salaneck, E., 2004. Molecular evolution of NPY receptor subtypes. Neuropeptides 38, 141-151.
Leary, P.J., et al., 2016. Histamine H2 Receptor Antagonists, Left Ventricular Morphology, and Heart Failure Risk: The MESA Study. J Am Coll Cardiol 67, 1544-1552.
Leguia, M., Wessel, G.M., 2006. The histamine H1 receptor activates the nitric oxide pathway at fertilization. Mol Reprod Dev 73, 1550-1563.
Leurs, R., et al., 2005. The histamine H3 receptor: from gene cloning to H3 receptor drugs. Nat Rev Drug Discov 4, 107-120.
Leurs, R., et al., 2000. H3 receptor gene is cloned at last. Trends Pharmacol Sci 21, 11-12.
Leurs, R., et al., 1995. Lysine200 located in the fifth transmembrane domain of the histamine H1 receptor interacts with histamine but not with all H1 agonists. Biochem Biophys Res Commun 214, 110-117.
Leurs, R., et al., 1994. Site-directed mutagenesis of the histamine H1-receptor reveals a selective interaction of asparagine207 with subclasses of H1-receptor agonists. Biochem Biophys Res Commun 201, 295-301.
Li, Q., et al., 2015. Non-classical amine recognition evolved in a large clade of olfactory receptors. Elife 4, e10441.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658-1659.
Liu, C., et al., 2001. Cloning and pharmacological characterization of a fourth histamine receptor (H(4)) expressed in bone marrow. Mol Pharmacol 59, 420-426.

Louis, A., et al., 2013. Genomicus: five genome browsers for comparative genomics in eukaryota. Nucleic Acids Res 41, D700-705.
Lovenberg, T.W., et al., 1999. Cloning and functional expression of the human histamine H3 receptor. Mol Pharmacol 55, 1101-1107.
Merchant, S., et al., 2014. Unexpected cross-species contamination in genome sequencing projects. PeerJ 2, e675.
Minh, B.Q., et al., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30, 1188-1195.
Misof, B., et al., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763-767.
Moguilevsky, N., et al., 1995. Pharmacological and functional characterisation of the wild-type and site-directed mutants of the human H1 histamine receptor stably expressed in CHO cells. J Recept Signal Transduct Res 15, 91-102.
Morisset, S., et al., 2000. High constitutive activity of native H3 receptors regulates histamine neurons in brain. Nature 408, 860-864.
Morris, G.M., et al., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30, 2785-2791.
Morse, K.L., et al., 2001. Cloning and characterization of a novel human histamine receptor. J Pharmacol Exp Ther 296, 1058-1066.
Munk, C., et al., 2016. GPCRdb: the G protein-coupled receptor database - an introduction. Br J Pharmacol 173, 2195-2207.
Nakamura, T., et al., 2000. Molecular cloning and characterization of a new human histamine receptor, HH4R. Biochem Biophys Res Commun 279, 615-620.
Nault, M.A., et al., 2002. Effects of the selective H1 and H2 histamine receptor antagonists loratadine and ranitidine on autonomic control of the heart. Anesthesiology 96, 336-341.
Nguyen, L.T., et al., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32, 268-274.
Nguyen, T., et al., 2001. Discovery of a novel member of the histamine receptor family. Mol Pharmacol 59, 427-433.
Ocampo Daza, D., Larhammar, D., 2018. Evolution of the growth hormone, prolactin, prolactin 2 and somatolactin family. Gen Comp Endocrinol 264, 94-112.
Oda, T., et al., 2000. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. J Biol Chem 275, 36781-36786.
Ohta, K., et al., 1994. Site-directed mutagenesis of the histamine H1 receptor: roles of aspartic acid107, asparagine198 and threonine194. Biochem Biophys Res Commun 203, 1096-1101.
Panula, P., et al., 2015. International Union of Basic and Clinical Pharmacology. XCVIII. Histamine Receptors. Pharmacol Rev 67, 601-655.
Pedersen, J.E., et al., 2018. Evolution of the Muscarinic Acetylcholine Receptors in Vertebrates. eNeuro 5.
Pedersen, J.E., et al., 2019. Evolution of vertebrate nicotinic acetylcholine receptors. BMC Evol Biol 19, 38.
Potnuri, A.G., et al., 2016. Targeting histamine-2 receptor for prevention of cardiac remodelling in chronic pressure overload. Int J Cardiol 202, 831-833.
Puttonen, H.A.J., et al., 2018. Knockout of histamine receptor H3 alters adaptation to sudden darkness and monoamine levels in the zebrafish. Acta Physiol (Oxf) 222.
Qiu, H., et al., 2011. Unresolved orthology and peculiar coding sequence properties of lamprey genes: the KCNA gene family as test case. BMC Genomics 12, 325.
Roeder, T., 2003. Metabotropic histamine receptors--nothing for invertebrates? Eur J Pharmacol 466, 85-90.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.
Ronquist, F., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61, 539-542.
Ruiz-Trillo, I., et al., 2008. A phylogenomic investigation into the origin of metazoa. Mol Biol Evol 25, 664-672.
Schneider, E.H., Seifert, R., 2016. The histamine H4-receptor and the central and peripheral nervous system: A critical analysis of the literature. Neuropharmacology 106, 116-128.
Sefideh, F.A., et al., 2014. Local duplication of gonadotropin-releasing hormone (GnRH)

receptor before two rounds of whole genome duplication and origin of the mammalian GnRH receptor. PLoS One 9, e87901.
Seifert, R., et al., 2013. Molecular and cellular analysis of human histamine receptor subtypes. Trends Pharmacol Sci 34, 33-58.
Shimamura, T., et al., 2011. Structure of the human histamine H1 receptor complex with doxepin. Nature 475, 65-70.
Shin, N., et al., 2002. Molecular modeling and site-specific mutagenesis of the histamine-binding site of the histamine H4 receptor. Mol Pharmacol 62, 38-47.
Simion, P., et al., 2017. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. Curr Biol 27, 958-967.
Smit, M.J., et al., 1996. Inverse agonism of histamine H2 antagonist accounts for upregulation of spontaneously active histamine H2 receptors. Proc Natl Acad Sci U S A 93, 6802-6807.
Smith, J.J., et al., 2009. Programmed loss of millions of base pairs from a vertebrate genome. Proc Natl Acad Sci U S A 106, 11212-11217.
Smith, J.J., et al., 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet 45, 415-421, 421e411-412.
Soubrier, J., et al., 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol Biol Evol 29, 3345-3358.
Spielman, S.J., et al., 2015. Comprehensive, structurally-informed alignment and phylogeny of vertebrate biogenic amine receptors. PeerJ 3, e773.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.
Steel, D.J., et al., 1997. Identification of an 8-lipoxygenase pathway in nervous tissue of Aplysia californica. J Biol Chem 272, 18673-18681.
Sutherby, J., et al., 2012. Histamine is a modulator of metamorphic competence in *Strongylocentrotus purpuratus* (Echinodermata: Echinoidea). BMC Dev Biol 12, 14.
Tardivel-Lacombe, J., et al., 2001. Chromosomal mapping and organization of the human histamine H3 receptor gene. Neuroreport 12, 321-324.
ter Laak, A.M., et al., 1995. Modelling and mutation studies on the histamine H1-receptor agonist binding site reveal different binding modes for H1-agonists: Asp116 (TM3) has a constitutive role in receptor stimulation. J Comput Aided Mol Des 9, 319-330.
Thurmond, R.L., et al., 2008. The role of histamine H1 and H4 receptors in allergic inflammation: the search for new antihistamines. Nat Rev Drug Discov 7, 41-53.
Tostivint, H., et al., 2014. Molecular evolution of GPCRs: Somatostatin/urotensin II receptors. J Mol Endocrinol 52, T61-86.
Trifinopoulos, J., et al., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44, W232-W235.
Uveges, A.J., et al., 2002. The role of transmembrane helix 5 in agonist binding to the human H3 receptor. J Pharmacol Exp Ther 301, 451-458.
Wernersson, R., Pedersen, A.G., 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res 31, 3537-3539.
West, R.E., et al., 1990. Identification of two H3-histamine receptor subtypes. Molecular Pharmacology 38, 610-613.
Wieland, K., et al., 1999. Mutational analysis of the antagonist-binding site of the histamine H(1) receptor. J Biol Chem 274, 29994-30000.
Yamada, K.D., et al., 2016. Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. Bioinformatics 32, 3246-3251.
Yamamoto, K., Vernier, P., 2011. The evolution of dopamine systems in chordates. Front Neuroanat 5, 21.
Yamashita, M., et al., 1991. Expression cloning of a cDNA encoding the bovine histamine H1 receptor. Proc Natl Acad Sci U S A 88, 11515-11519.
Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. Journal of Molecular Evolution 39, 306-314.

Yang, Z., 1995. A space-time process model for the evolution of DNA sequences. Genetics 139, 993-1005.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24, 1586-1591.
Yang, Z., et al., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22, 1107-1118.
Yoshihisa, A., et al., 2017. Associations of Acid Suppressive Therapy With Cardiac Mortality in Heart Failure Patients. J Am Heart Assoc 6.
Yun, S., et al., 2015. Prevertebrate Local Gene Duplication Facilitated Expansion of the Neuropeptide GPCR Superfamily. Mol Biol Evol 32, 2803-2817.
Zamanian, M., et al., 2011. The repertoire of G protein-coupled receptors in the human parasite *Schistosoma mansoni* and the model organism *Schmidtea mediterranea*. BMC Genomics 12, 596.
Zeng, Z., et al., 2014. Disruption of histamine H2 receptor slows heart failure progression through reducing myocardial apoptosis and fibrosis. Clin Sci (Lond) 127, 435-448.
Zhang, M., et al., 2007. The histamine H(4) receptor: a novel modulator of inflammatory and immune disorders. Pharmacol Ther 113, 594-606.
Zhu, Y., et al., 2001. Cloning, expression, and pharmacological characterization of a novel human histamine receptor. Mol Pharmacol 59, 434-441.

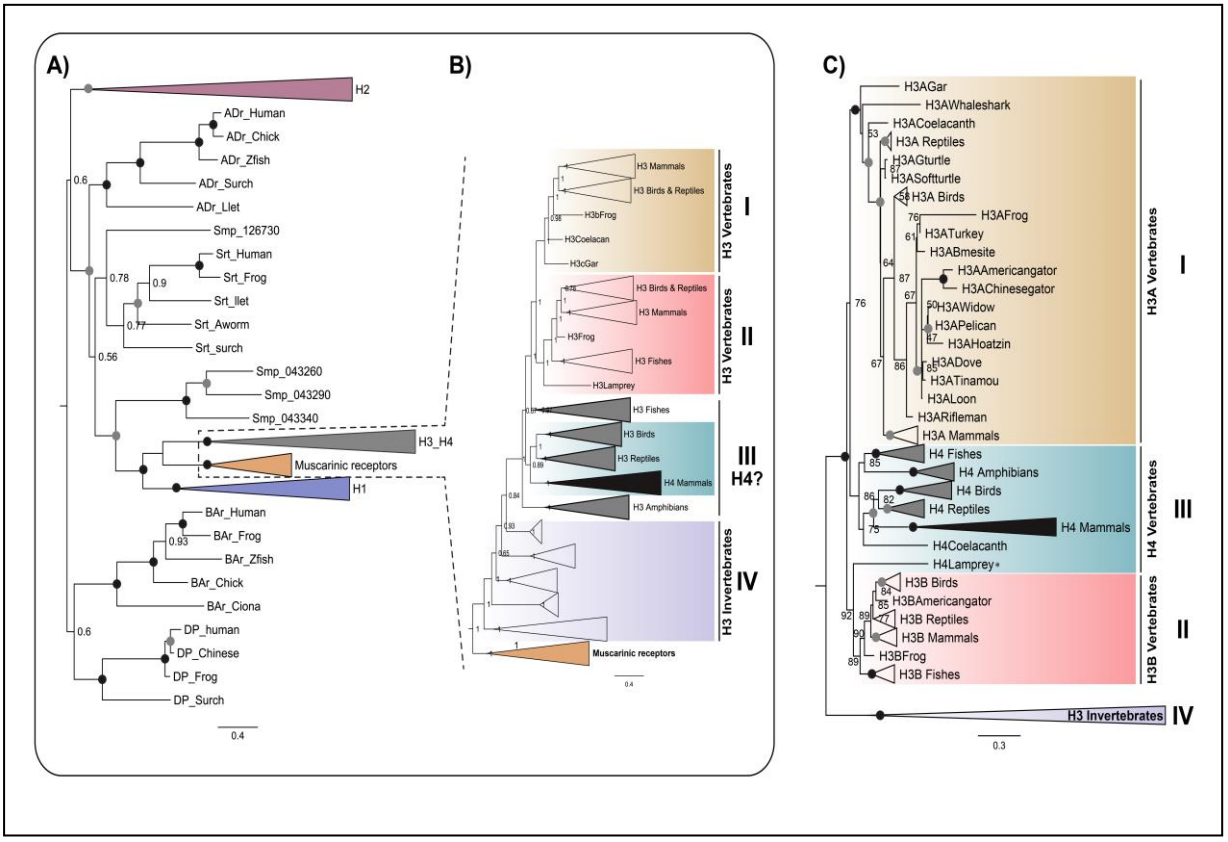

Figure 1. Phylogenetic analyses of histamine receptors. A) Phylogenetic relationship of histamine, adrenoceptor, muscarinic acetylcholine, dopamine and serotonin receptors. The flatworm specific histamine receptors are labeled Smp_043260, Smp_043290 and Smp_043340. Adrenoceptors α- (ADr) and β-(BAr), serotonin receptors (Srt), muscarinic acetylcholine receptors and dopamine receptors (DP) were used as outgroups. The node containing BAr and DP was used as a mid-point root for better visualization. The tree file in Newick format is given in Tree S2a. The tree was generated using the Bayesian approach implemented in MrBayes 3.2.6 using JTT+F+G model based on ProtTest v3.3. Numbers on the nodes represents the Bayesian posterior probabilities (BPP). **B)** Expansion of the collapsed node containing the H_3-H_4 and muscarinic receptors. The H_3-H_4 sequences cluster into four major clades labeled I, II, III and IV. The subtype annotations indicated here are as present in current databases. The H_3 clades are paraphyletic and clade III displays an intriguing placement of the H_3 and H_4 clades and is marked with H4 with **?** symbol. **C)** Phylogenetic relationships within the H_3A, H_3B and H_4 subtypes incorporating the revised nomenclature. The revised annotation of subtypes resulted from an analysis of phylogenetic data in combination with synteny of gene-neighborhoods. The tree file in Newick format is given in Tree S2b. The tree was generated using ML method implemented in IQ-TREE using JTT+R6 model based on ModelFinder. Numbers on the nodes represents the percentage bootstrap support (BS). Some nodes were compressed for ease of representation. Black circle represents 100% BPP/BS and grey circle represents >95% BPP/BS. Scale bar represents the number of estimated changes per site for a unit branch length. The receptor group abbreviations, names and accession numbers of the sequences and common and binomial names of the species are listed in Table S1. The H4Lamprey sequence marked with * symbol was named based on gene synteny data.

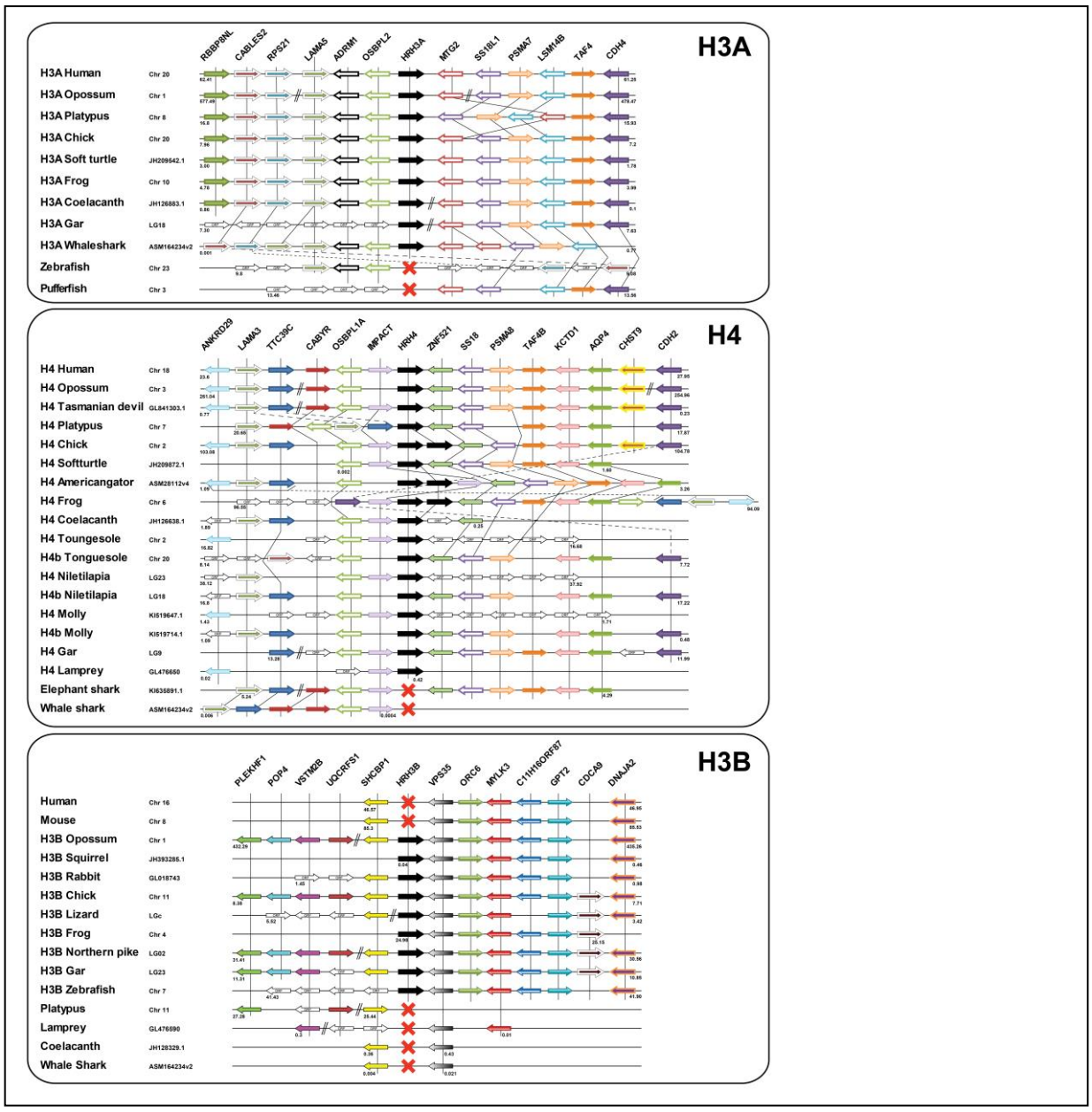

Figure 2. Gene neighborhood analysis of H_3-H_4 receptor subtypes. Patterns of conserved gene neighborhood of H_3A, H_3B and H_4 in their chromosomal regions. H_3A and H_4 share conserved gene neighborhood indicating their common ancestry. H_4 paralogons confirms their presence across vertebrates. The H_3B subtype paralogon is distinct from that of the H_3A and H_4.

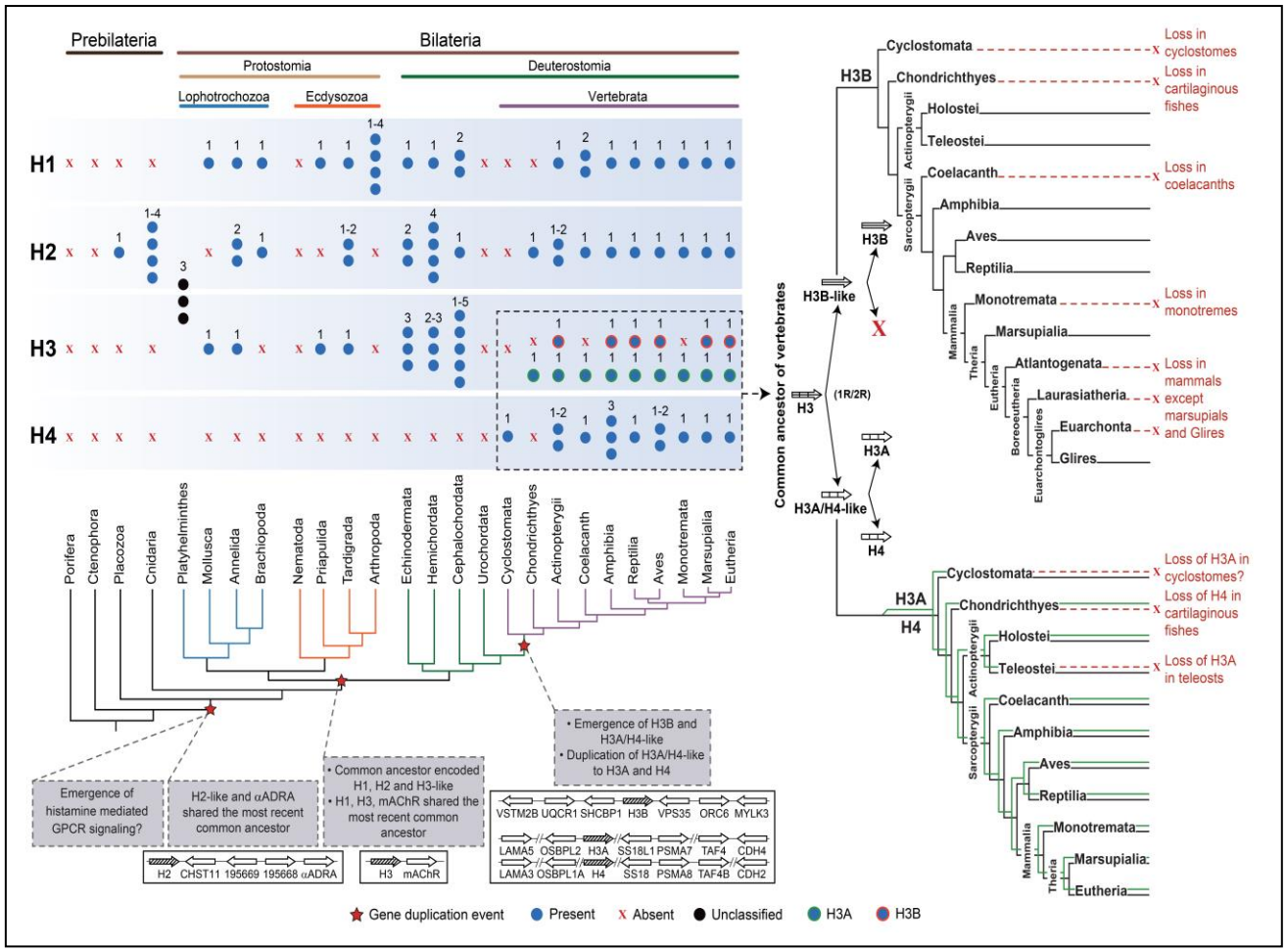

Figure 3. Phyletic distribution and proposed evolutionary history of histamine receptor subtypes in the Metazoa. The presence and absence of each subtype are marked with respect to each phylum, class and order. The number of duplicates in each subtype is given. The metazoan tree topology is adapted from (Koutsovoulos et al., 2016; Merchant et al., 2014; Ruiz-Trillo et al., 2008; Simion et al., 2017). Branches on the schematic tree do not represent actual evolutionary distances. The model on the right presents the evolution of the H_3A, H_3B and H_4.

Appendix A. Supplementary material
Supplementary information includes:
Supplementary note
Supplementary figures S1-S9
Supplementary tables S1-S12
Supplementary data S1-S3
Supplementary references