Molecular Evolution of the Transmembrane Domains of G Protein-Coupled Receptors

Sarosh N. Fatakia, Stefano Costanzi, Carson C. Chow*

Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, United States of America

Abstract

G protein-coupled receptors (GPCRs) are a superfamily of integral membrane proteins vital for signaling and are important targets for pharmaceutical intervention in humans. Previously, we identified a group of ten amino acid positions (called key positions), within the seven transmembrane domain (7TM) interhelical region, which had high mutual information with each other and many other positions in the 7TM. Here, we estimated the evolutionary selection pressure at those key positions. We found that the key positions of receptors for small molecule natural ligands were under strong negative selection. Receptors naturally activated by lipids had weaker negative selection in general when compared to small molecule-activated receptors. Selection pressure varied widely in peptide-activated receptors. We used this observation to predict that a subgroup of orphan GPCRs not under strong selection may not possess a natural small-molecule ligand. In the subgroup of MRGX1-type GPCRs, we identified a key position, along with two non-key positions, under statistically significant positive selection.

Introduction

G protein-coupled receptors (GPCRs) constitute a diverse superfamily of integral membrane proteins involved in intercellular signal transduction. Their genes are expressed in almost all eukaryotes [1,2,3,4,5]. The receptor consists of a single polypeptide chain that loops through the cell membrane seven times to form an interhelical cavity of seven alpha-helical transmembrane domains (7TMs). GPCRs are the largest superfamily of integral membrane proteins in humans. About half of the GPCRs in the human genome are non-olfactory receptors [6,7,8]. These receptors mediate vital physiological functions and are a major target for pharmaceutical interventions [9,10]. Although diverse in sequence composition and function, GPCRs share a common molecular architecture of 7TMs connected via three intracellular and three extracellular loops. Fredriksson and Schioth have categorized the GPCRs into five distinct families [8,11] - Glutamate (also known as class C), Rhodopsin (also known as class A), Adhesion, Secretin (collectively known as class B) and Frizzled/Taste (also known as class F). Nearly 85% of the non-olfactory receptors belong to class A. Class A receptors bind different natural ligands that range from small-molecules such as ADP to larger ones such as neuropeptides or chemokines.

New protein functions in paralogous protein superfamilies arise by the modulation of older existing ones [12]. During this evolutionary process, some of the amino acid residues remain conserved. However, mutations of some residues may be followed by compensatory mutations elsewhere to preserve function or give rise to new ones. The identification of such related residue positions can help to identify biologically relevant sets of residues in protein superfamilies. Previously, we identified a set of positions in the interhelical cavity enclosed within the 7TM domain of class A GPCRs that have high mutual information (MI) with other positions and each other [13,14]. These key positions were found to be located in the region that constitutes the binding cavity of GPCRs whose structures have been solved. Biochemical data suggest that this region hosts the orthosteric binding cavity for all class A GPCRs naturally activated by small molecules.

Here, we examine the nucleotide sequences corresponding to these GPCRs to probe the evolutionary selection pressure at these key positions. Synonymous nucleotide substitutions (‘silent’ mutations) do not change the translated amino acid sequence so their substitution rate $d_S$ (also referred to as $K_S$) is not subject to selective pressure on the expressed protein. Nonsynonymous mutations alter the amino acid sequence and their substitution rate $d_N$ (also referred to as $K_N$) is a function of selective pressure on the protein. The ratio $d_S/d_N$, referred to as $\omega$, gives a measure of the selection pressure at that site [15,16]. When there exists negative or purifying selection pressure at a codon position, $\omega<1$ and synonymous substitutions dominate. When the position is under positive or adaptive selection, $\omega>1$ and nonsynonymous substitutions dominate. Rare instances of positive selection are of special interest in tracing functional divergence among protein families and physiological adaptations in humans [17,18,19,20,21]. When the position evolves neutrally – without any strong preferential selection, the two substitution rates are nearly equal. Here we determine $\omega$ at the key positions and compare it to other 7TM positions. If the selection pressure at the key positions is less neutral then on other positions then this supports the hypothesis that the
### Table 1. List of class A GPCRs included in the study.

| Subgrp idx | # GPCRs in subgrp | GPCRs included in the subgroups<sup>a</sup> | Natural ligand | Chemical class of natural ligand<sup>b</sup> | Notes<sup>c</sup> |
|------------|-------------------|---------------------------------|----------------|---------------------------------|-----------------|
| 1          | 5                 | CHRM1 (ACM1), CHRM2 (ACM2), CHRM3 (ACM3), CHRM4 (ACM4), CHRM5 (ACM5) | acetylcholine | small |                               |
| 2          | 5                 | DRD1, DRD2, DRD3, DRD4, DRD5 | dopamine | small |                               |
| 3          | 5                 | P2RY12 (P2Y12), P2RY13 (P2Y13), P2RY14 (P2Y14), GPR87, GPR171<sup>GP171</sup> | nucleotides, lysophosphatidic acid (GPR87) | small | o, S                           |
| 4          | 7                 | HTR1A (SHT1A), HTR1B (SHT1B), HTR1D (SHT1D), HTR1F (SHT1F), HTR1E (SHT1E), HTR5A (SHT5A), HTR7 (SHT7R) | 5-hydroxytryptamine | small |                               |
| 5          | 5                 | P2RY1, P2RY2, P2RY4, P2RY6, P2RY11 (P2Y11) | nucleotides | small |                               |
| 6          | 3                 | MTNR1A (MTR1A), MTNR1B (MTR1B), GPR50<sup>MTR1L</sup> | melatonin | small | o                             |
| 7          | 5                 | ADRA1A (ADA1A), ADRA1B (ADA1B), ADRB1, ADRB2, ADRB3 | adrenaline | small |                               |
| 8          | 3                 | HTR2A (SHT2A), HTR2B (SHT2B), HTR2C (SHT2C) | 5-hydroxytryptamine | small |                               |
| 9          | 4                 | HRH1, HRH2, HRH3, HRH4 | histamine | small |                               |
| 10         | 3                 | ADORA1 (AA1R), ADORA2A (AA2AR), ADORA2B (AA2BR) | adenosine | small |                               |

<sup>a</sup>The receptors are indicated through their gene name. Uniprot name, when different from the gene name, and common synonyms are listed in parentheses. Orphan receptors are indicated in bold and indicated as ‘o’ in Notes.

<sup>b</sup>Small indicates “small molecules” and refers to biogenic amines, nucleosides and nucleotides.

<sup>c</sup>The symbol “o” indicates that the subgroup has one or more orphan GPCR.

doi:10.1371/journal.pone.0027813.t001

### Table 2. List of class A GPCRs included in the study (continued from Table 1).

| Subgrp idx | # GPCRs in subgrp | GPCRs included in the subgroups<sup>a</sup> | Natural ligand | Chemical class of natural ligand | Notes<sup>c</sup> |
|------------|-------------------|---------------------------------|----------------|---------------------------------|-----------------|
| 11         | 6                 | S1PR2<sup>EDG5</sup>, S1PR1<sup>EDG1</sup>, S1PR3<sup>EDG3</sup>, S1PR5<sup>EDG8</sup>, LPAR1<sup>EDG2</sup>, LPAR3<sup>EDG7</sup> | sphingosine 1-phosphate, lysophosphatidic acid (LPAR1, LPAR3) | lipid |                               |
| 12         | 3                 | GPR3, GPR6, GPR12 | sphingosine 1-phosphate | lipid |                               |
| 13         | 3                 | FFAR1<sup>GPR40</sup>, FFAR2<sup>GPR43</sup>, FFAR3<sup>GPR41</sup> | free fatty acids | lipid |                               |
| 14         | 7                 | PTGDR<sup>PD2R</sup>, PTGER1<sup>PE2R1</sup>, PTGER3<sup>PE2R3</sup>, PTGER4<sup>PE2R4</sup>, PTGFR<sup>PF2R</sup>, PTGIR<sup>PZ2R</sup>, TXB2A<sup>TZ2R</sup> | prostaglandins, thromboxane (TA2R) | lipid |                               |
| 15         | 3                 | CYSLTR1<sup>CLTR1</sup>, CYSLTR2<sup>CLTR2</sup>, GPR17 | cysteinyl leukotrienes | lipid |                               |
| 13b<sup>d</sup> | 4                 | FFAR1<sup>GPR40</sup>, FFAR2<sup>GPR43</sup>, FFAR3<sup>GPR41</sup>, GPR42 (pseudogene) | free fatty acids | lipid |                               |
| 16         | 5                 | LPAR4<sup>P2Y9</sup>, LPAR6<sup>P2Y5</sup>, GPR174<sup>GP174</sup>, P2RY10<sup>P2Y10</sup>, PTAFR | lysophosphatidic acid, sphingosine 1-phosphate (P2Y10), platelet activating factor | lipid | o                             |
| 17         | 5                 | RRH<sup>OPPSX</sup>, OPN3, OPN4, OPNS, RGR | retinoids | lipid |                               |
| 18         | 4                 | OPN1MW<sup>OPSG</sup>, OPN1LW<sup>OPSR</sup>, RHO<sup>OPSD</sup>, OPN1SW<sup>OPSB</sup> | retinoids | lipid |                               |
| 19         | 3                 | GPR81, GPR109B<sup>G109B</sup>, GPR109A<sup>G109A</sup> | hydroxylated short and medium-chain fatty acids | lipid |                               |

<sup>a</sup>The receptors are indicated through their gene name. Uniprot name, when different from the gene name, and common synonyms are listed in parentheses. Orphan receptors are indicated in bold and indicated as ‘o’ in Notes.

<sup>b</sup>The symbol “o” indicates that the subgroup has one or more orphan GPCR.

<sup>d</sup>Derived from subgroup 13 through the addition of the pseudogene GPR42.

doi:10.1371/journal.pone.0027813.t002
high mutual information between the key positions and associated high entropy did not simply arise from evolutionary drift.

Results

All subgroups of human GPCRs were classified into three categories in terms of their natural ligands: 1) small molecules (including biogenic amines, nucleosides and nucleotides), 2) lipids, and 3) peptides. GPCR subgroups whose natural ligands could not be exclusively classified as any of the above were categorized as divergent. A number of human GPCRs are orphans with no known natural ligands. The list of GPCR subgroups and the chemical class of associated natural ligands is in Tables 1, 2, 3 and 4. Of the 45 subgroups of GPCRs, excluding subgroup 13b, 10 subgroups are activated by small molecules listed in Table 1, 9 subgroups are activated by lipids listed in Table 2, and 19 subgroups are activated by peptides listed in Table 3. Six subgroups were categorized as divergent, because they are activated by natural ligands that belong to different chemical classes or contain two or more orphans. One subgroup exclusively contained human orphan GPCRs. The divergent and orphan subgroups are listed in Table 4.

The $\omega$ values were determined for subgroups with at least three paralogs. Selection pressure at the key positions, $\omega_{key}$, is shown in Figure 1. The $\omega_{key}$ and its average, $<\omega_{key}>$, of subgroups associated with small molecules differed from that of subgroups associated with lipids and peptides. The Kruskal-Wallis rank sum test showed that $<\omega_{key}>$ for small molecule-activated receptors had significantly lower values compared to subgroups of lipid-activated receptors, peptide-activated receptors and divergent receptors ($p<0.003$). The $\omega_{key}$ values from all ten subgroups activated by small molecules showed strong negative selection ($\omega<0.05$).

We confirmed that human MRGX1-type receptors are under positive selection [22,23]. Positive selection at three positions was inferred in subgroup 30 (MRGX1, MRGX2, MRGX3 and MRGX4 pain receptors) using three different tests. The results of the likelihood ratio estimates are shown in Table 5. The results of $\omega$ for key positions and positions with posterior probability of positive selection exceeding 0.5 are shown in Table 6. We inferred strong positive selection at key position 3.29 in the Ballesteros-Weinstein scheme [24], ($\omega = 6.3$, posterior probability for $\omega>1 = 0.998$). Two non-key positions: 2.56 ($\omega = 6.1$, posterior probability for $\omega>1 = 0.948$) and 2.60 ($\omega = 6.1$, posterior probability for $\omega>1 = 0.947$) were also under positive selection. Six of the key positions (3.35, 3.33, 5.42, 6.55, 7.35 and 7.39) were not under statistically significant positive selection. Three key positions (3.32, 4.60 and 5.39) were under negative selection. Subgroup 41 (MAS1L, MRGRD, MAS, and MRGRF pain receptors) did not show any statistically significant signature for positive selection. Previous studies had demonstrated positive selection pressure for the combined subgroups 41 and 30 [MRG receptors from humans

Table 3. List of class A GPCRs included in the study (continued from Tables 1, 2).

| Subgrp idx | # GPCRs in subgrp | GPCRs included in the subgroups* | Natural ligand | Chemical class of natural ligandb | Notesc |
|------------|-------------------|----------------------------------|----------------|----------------------------------|--------|
| 20         | 3                 | TACR1 (NK1R), TACR1 (NK2R), TACR3 (NK3R) | tachykinin neuropeptides | peptide |         |
| 21         | 3                 | TSHR, LHCGR (LSHR), FSHR | glycoprotein hormones | peptide |         |
| 22         | 4                 | F3R (PAR1), F2RL1 (PAR2), F2RL2 (PAR3), F2RL3 (PAR4) | unmasked N-terminus | peptide |         |
| 23         | 5                 | GPR83, NPY1R, NPY2R, PYY1 (NPY4R), NPY5R | neuropeptide Y and peptide YY | peptide | o |
| 24         | 3                 | C3AR1 (C3AR), C5AR1 (C5AR), GPR77 (C5ARL) | anaphylatoxins | peptide |         |
| 25         | 4                 | EDNRA, EDNRB, GPR37, GPR37L1 (ETRB2) | Endothelins | peptide | o |
| 26         | 5                 | LGR5, LGR6, RXFP1 (LGR7), RXFP2 (LGR8) | Relaxin | peptide |         |
| 27         | 3                 | GALR1, GALR2, GALR3 | Galanin | peptide | N |
| 28         | 4                 | OPR1L (OPRX), OPRM1 (OPRM), OPR1D (OPRD), OPRK1 (OPRK) | opioid peptides | peptide | N |
| 29         | 3                 | SSTR2 (SSR2), SSTR3 (SSR3), SSTR5 (SSR5) | somatostatins | peptide |         |
| 30         | 3                 | GRPR, NMBR, BR53 | bombesin-related peptides | peptide |         |
| 31         | 3                 | MC3R, MC4R, MCSR | melanocortins | peptide | N |
| 32         | 3                 | AVPR1A (V1AR), AVPR1B (V1BR), AVPR2 (V2R) | Vasopressin | peptide | N |
| 33         | 10                | CXC1R, CXC2R, CXC3R, CXC4, CXC5, CXC6, CCR6, CCR7, CCR9, CCR10 | Chemokines | peptide |         |
| 34         | 5                 | APLNR (APJ), AGTR1 (AG2R, AG2S), RL3R1 (RLN3R2), RXFP4 (RLN3R2) | apelin (APLNR), angiotensin (AGTR1), relaxin (RL3R1, RLN3R2) | peptide |         |
| 35         | 3                 | NT5R1 (NTR1), NT5R2 (NTR2), GPR39 | neurotensin, obestatin (GPR39) | peptide |         |
| 36         | 9                 | CCR1, CCR2, CCR3, CCR4, CCR5, CCR8, CCR2L, CXC3R (CXC3R, C3X1), CCBP2 | Chemokines | peptide | S |
| 37         | 3                 | FPR1, FPR2 (FPRL1), FPR3 (FPRL2) | N-formyl-methionyl peptides (FPRs) | peptide |         |
| 38         | 4                 | MRGPRX1 (MRGX1), MRGPRX2 (MRGX2), MRGPRX3 (MRGX3), MRGPRX4 (MRGX4) | enkephalins (MRGPRX1), cortistatins (MRGPRX2) | peptide | o |

*The receptors are indicated through their gene name. Uniprot name, when different from the gene name, and common synonyms are listed in parentheses. Orphan receptors are indicated in bold and indicated as ‘o’ in Notes.

*The Symbol “N” indicates that pairs of receptors of the subgroup do not satisfy max$(d_3)<1$ in the Nei-Gojobori counting scheme [64]. The symbol “S” indicates that pairs of receptors from the subgroup do not satisfy max $(d_3)<3$ in the Nei-Gojobori scheme. The symbol “o” indicates that the subgroup has one or more orphan GPCR.

doi:10.1371/journal.pone.0027813.t003
and model organisms) [22,23]. We inferred that the combined subgroup, 44, also exhibited positive selection exclusively within 7TMs but subgroup 41 did not exclusively exhibit statistically significant positive selection. Results from the likelihood ratio test for subgroup 44 are included in Table 5. An independent analysis of subgroup 44 confirmed statistically significant positive selection at key positions 3.29 and 5.35 along with two non-key positions 2.57 and 2.60. (Position 2.60 showed positive selection in subgroup 38 but not position 2.57).

We next compared $\theta_{\text{key}}$ to random sets of 7TM positions $\theta_{\text{random7TM}}$ to see if there was stronger selection pressure at the key positions. The values are shown in Figure 2 and Figure S1. For most receptor subgroups binding to small molecules, $\theta_{\text{key}}$ was less than $\theta_{\text{random7TM}}$ although within two standard deviations of $\theta_{\text{random7TM}}$. The selection pressure for subgroup 42 was atypical in that $\theta_{\text{key}}$ was larger than $\theta_{\text{random7TM}}$ by two standard deviations. For six of nine subgroups associated with lipid-activated receptors, $\theta_{\text{key}}$ was nearly equal to $\theta_{\text{random7TM}}$. In subgroups activated by peptides, $\theta_{\text{key}}$ was less than or nearly equal to $\theta_{\text{random7TM}}$. Subgroup 38, which exhibits strong positive selection, was the only case where $\theta_{\text{key}}$ exceeded $\theta_{\text{random7TM}}$ by two standard deviations. Linear regression of $\theta_{\text{key}}$ vs. $\theta_{\text{random7TM}}$ for the subgroups excluding subgroup 38 and 44, showed a linear dependence ($R^2 = 0.892$, $p<2.2 \times 10^{-16}$) (See Figure S2). However, as seen in Figure 3, $\theta_{\text{key}}$ vs. $\theta_{\text{random7TM}}$ is less than unity for small $\theta_{\text{key}}$, and increases significantly with $\theta_{\text{key}}$, ($p<3.6 \times 10^{-6}$) and $\theta_{\text{random7TM}}$ ($p<4.9 \times 10^{-7}$). The dependence remained significant even after including subgroup 38. The Yang and Swanson’s “fixed sites” model [23] indicated that $\theta_{\text{key}}$ was significantly lower than $\theta_{\text{random7TM}}$ in two of the ten small molecule subgroups (subgroups 3 and 10). Subgroup 11, which consists of lipid-activated receptors, showed statistically significant differences between key and random positions. In 5 of the 19 subgroups of the peptide receptors, key positions have significantly higher selection pressure than random positions. Only subgroup 22 of the peptide-activated receptors was significantly lower. The results are summarized in Table S1.

We also tested if the diversity of $\theta_{\text{key}}$ values in subgroups was due to the dissimilarity among amino acid (AA) residues at a given MSA position since it is expected that stronger selection pressure should result in lower variability. However, the strength of the correlation between $\theta_{\text{key}}$ and variability was not known. We examined this with three different measures. First, we computed the Shannon entropy ($H$) for the key positions of each subgroup, which has a theoretical range of 0 bits $\leq H \leq 4.32$ bits. Figure S3 shows $H$ for every key position across all subgroups. Figure S4 is a plot of $H$ vs. $\theta_{\text{key}}$ for subgroups with average pair-wise max($\theta_{\text{key}}$) < 1 (see Materials and Methods). This figure shows a slight trend of higher entropy for higher $\theta_{\text{key}}$, although it was not statistically significant. A linear regression of $H_{\text{key}}$ against log$_{10}$($\theta_{\text{key}}$) found a correlation coefficient of $R = 0.47$ ($p<1.4 \times 10^{-8}$). However, the regression of $H_{\text{key}}$ against log$_{10}$($\theta_{\text{key}}$) had much lower correlation when $\theta_{\text{key}}$ restricted to $\theta_{\text{key}} < 0.1$ ($R = 0.26$, $p<9.8 \times 10^{-5}$). However, this decrease in correlation could be due to the decrease in statistical power because the sample size is reduced. Similar results were found using the BLOSUM60 substitution matrix [26] and a distance matrix $D_{\text{key}}$ to estimate the dissimilarity among residues within subgroups at key positions. Results are in Figures S3, S6, S7, and S8. These results show that AA variability at MSA positions is only weakly correlated with $\theta_{\text{key}}$ and the correlation is weaker for subgroups under strong negative selection.

Table 4. List of class A GPCRs included in the study (continued from Tables 1, 2 and 3).

| Subgrp idx | # GPCRs in subgroup | GPCRs included in the subgroups* | Natural ligand | Chemical class of natural ligand* | Notes* |
|------------|---------------------|---------------------------------|---------------|---------------------------------|--------|
| 39         | 5                   | GPR101 (GPR101), GPR161 (G1P161), GPR133 (GPR135), GPR63, GPR45 | sphingosine 1-phosphate | divergent | o |
| 40         | 3                   | GPR4, GPR65 (PSYR), GPR68 (OGR1) | protons (GPR4 and GPR68), glycosphingolipids (GPR65) | divergent | N |
| 41         | 4                   | MAS1 (MAS5), MAS1L (MARG), MPRGPRD (MARGDR), MPRGPRF (MARGFR), GPR140 | angiotensin (MAS1), [l-alanine (MARGRD)) | divergent | o |
| 42         | 5                   | TAA1 (TAR01), TAA5, TAA6 (TAR4), TAA8 (TAR5), TAA9 (TAR3) | trace amines | divergent | o,S |
| 43         | 10                  | C3AR1 (C3AR1), MRGPRD (C3AR1), CMLR1 (CML1), FPR1, FPR2 (PPR1), FPR3 (PPR2), GPR1, GPR32, GPR44 (CATH2) | Anaphylatoxins (C3AR1, C3AR1, GPR77), chemokines (CMLR1), N-formyl-methionyl peptides (PPRs), chemerin (GPR1), resolvins (GPR32), prostanoids (GPR44) | divergent | d |
| 44         | 8                   | MAS1 (MAS5), MARG, MPRGPRD (MARGRD), MPRGPRF (MARGFR), GPR140, MRGPRX1 (MRGX1), MRGPRX2 (MRGX2), MRGPRX3 (MRGX3), MRGPRX4 (MRGX4) | angiotensin (MAS1), [l-alanine (MARGRD), enkephalins (MRGPRX1), cortistatins (MRGPRX2) | divergent | o |
| 45         | 3                   | GPR27, GPR85, GPR173 | orphans | o |
Molecular Evolution of GPCRs

Discussion

We have found that class A GPCR subgroups that are naturally activated by small molecules possessed strong negative selection in the key positions. Additionally, the selection pressure at the key positions is more likely to be stronger than the rest of the TM positions in small molecule receptors. The existence of strong negative selection supports coevolution over evolutionary drift as an explanation for the high mutual information between the key positions. We suggest that collective substitutions of key residues under strong selection pressure may have altered function in GPCRs. It has been shown previously that evolutionary characteristics such as phylogeny and sequence similarity of AA residues are a strong predictor of determinants of ligand specificity [27,28,29].

Under the rules of formal logic, the observation that small molecule receptors are always under strong negative selection at key positions allows for the prediction that GPCRs not under strong negative selection pressure are not naturally activated by small molecules. Based on our results from Figures 2 and S1, a threshold of \( \omega = 0.1 \) can be set to distinguish for strong negative selection (Figures 2 and S1 show that \( \omega_{\text{max}} = 0.05 \) and \( \omega_{\text{max}} = 0.1 \)). Thus, we can predict that receptor subgroups with \( \omega > 0.1 \) at the key positions do not possess a natural small molecule ligand. This would include orphan receptors MAS1L, MRGRPX, group 41, MRGRPX3, MRGRPX4 of 38 and 44, and TAAR5, TAAR6, TAAR8, and TAAR9 of 42. The inclusion of subgroup 42 may be considered to be surprising because TAAR1 of the group binds b-phenylethylamine and p-tryamine, which is a small molecule trace amine. Although this subgroup exhibits negative selection in conformation of recent studies involving TAAR orthologs [30,31] it is not strongly negative. This may imply that even though TAAR1 binds a trace amine, the key positions may not be vigorously maintaining their functionality.

Positive selection can lead to adaptation of a previous function [32,33,34,35]. Strong statistical evidence for positive selection was identified at key position 3.29 of subgroup 38 but not for subgroup 41, both of which are composed of MAS-related GPCRs. Statistical evidence for positive selection at key position 3.29 was identified in subgroup 44, with decreased statistical significance (results not shown). Because subgroup 44 comprises of subgroups 41 (MAS1L, MRGD, MAS, MRGRF) and 38 (MRGX1, MRGX2, MRGX3, MRGX4), sustained positive selection at 3.29 suggests adaptation specific to subgroup 38. Notably, in the 3D crystal structure of bovine rhodopsin [36], positions 3.29, 2.56 and 2.60 are near neighbors when represented on the resolved crystal structure of bovine rhodopsin in Figure 4. This suggests that, if there has been any novel or adaptive function in this interhelical cavity of MRGX1-type receptors, then it may have evolved via mutations (substitutions) that occurred in that circumscribed region of the receptor. Therefore, as a continuation of our novel bioinformatic approach, we identified an AA position from a cohort of statistically related AA positions in a protein family (namely, class A GPCRs) that evolves under strong positive selective pressure in a subgroup (namely, subgroup 38).

We examined entropy and measures of sequence similarity to test the hypothesis that strong selection pressure is related to low variability. Our results showed that even under strong negative selection pressure, sequence diversity remained. The wide diversity in selection pressure for receptors associated with the different classes of natural ligands was not attributable to the size of the subgroup. Diversity of \( \omega \) values is well documented [37,38,39,40] and for the different subgroups of GPCRs may be attributed to differences in the (i) natural ligands they bind, (ii) molecular mechanism of activation, (iii) phylogeny of the subgroups, and (iv) ubiquity of expression on cell surfaces [41,42,43].

The inclusion of orthologs would improve the accuracy of our analysis. We used three overlapping subgroups: 13b (overlapping with 13), 43 (overlapping with 37) and 44 (overlapping with 38 and 41) to probe how \( \omega_{\text{max}} \) and \( \omega_{\text{max}} \) changed with subgroup size. Subgroup 13b contained a pseudogene GPR42. Studies of class A GPCR orthologs have been previously investigated using opsins, MAS-related receptors, P2Y receptors and melanocortin receptors [22,23,44,45,46,47,48,49,50,51]. Amongst the GPCRs we studied, statistically significant positive selection has been widely reported for visual opsin receptors (receptors for trichromatic vision in old world primates) and subgroup 38 of MAS-related receptors (receptors for pain and itch). The divergence among human GPCR subgroups is varied and high polymorphism may be seen from recent studies, e.g. in the case of human MRGX1 receptors [52].

Table 5. P-value and likelihood ratio (LR) estimates from three PAML strategies for subgroups 38 and 44.

| PAML nested model pairs | subgroup 38 | subgroup 44 |
|-------------------------|-------------|-------------|
| Test 1 (M2a vs. M1a)    | 8.90        | 7.71        |
| Test 2 (M8 vs. M7)      | 8.94        | 17.65       |
| Test 3 (B vs. A)        | 7.98        | 31.82       |

Result of \( \Delta \) and P-value from Tests 1, 2 and 3. LR = 2\( \Delta = 2\ln\left(\frac{L_{\text{null}}}{L_{\text{null}}}\right) = 2\left(\ln L_{\text{null}} - \ln L_{\text{null}}\right)\). 

doi:10.1371/journal.pone.0027813.t005
Materials and Methods

Identification of key positions

An alignment of human non-olfactory class A 7TMs was obtained from [53]. Using that MSA, we identified a clique of statistically related MSA positions. These key positions had the highest collective MI with respect to one another and most other positions in the MSA [13,14]. The Ballesteros-Weinstein indexing scheme for GPCRs [24] was used to label all positions of the MSA.

Input data – nucleotide sequence data corresponding to 7TMs

Nucleotide sequence fragments that encoded the GPCR 7TMs were obtained from NCBI’s nucleotide database [54]. The cDNA sequence records encoding the entire protein sequence was extracted using NCBI’s Open Reading Frame online resource [55]. Entire AA sequence records were obtained from the RefSeq database [56] and the Uniprot database [57]. The amino acid and nucleotide sequence fragments from the 7TMs were concatenated. We used the IUPHAR 7TM receptor database [58,59] as well as a comprehensive GPCR listing from Gloriam et al. [60] to confirm our sequence data.

Input data – Phylogenetic tree

We used AA sequence fragments for the 7TMs of class A GPCRs to reconstruct a nearest neighbor phylogenetic tree. Program PROTDIST of PHYLIP [61] was used to compute phylogenetic distance across pairs of concatenated 7TM fragments using the JTT matrix for AA substitutions [62]. The nearest neighbor joining method [63] implemented in PHYLIP’s program NEIGHBOR was used to reconstruct the tree. Subgroups of GPCRs representing closely related 7TMs were identified from the phylogenetic tree, using a bootstrap approach. The selection of subgroup was refined using $d_N$ and $d_S$ selection criteria described below. A consensus phylogenetic tree was obtained using the CONSENSE program of PHYLIP. A list of GPCRs for all subgroups is shown in Tables 1, 2, 3 and 4.

GPCR subgroups

We analyzed forty-five subgroups, of which forty-two were non-overlapping and distinct. The number of constituent GPCRs in respective subgroups ranged from three to ten. Because GPCRs are highly divergent, we restricted the average maximum $d_N$ and maximum $d_S$ estimated from all pairs of receptors within subgroups unlike in a traditional analysis where subgroups may

---

**Table 6.** $\omega$ for subgroup 38.

| 7TM MSA position index | key | Ballesteros-Weinstein index | posterior probability ($\omega$>1) | NEB $\omega$ | comment |
|------------------------|-----|-----------------------------|-----------------------------------|--------------|---------|
| 8                      |     | 1.37                        | 0.562                             | 3.953        | -       |
| 19                     |     | 1.48                        | 0.680                             | 4.598        | -       |
| 49                     |     | 2.56                        | 0.948                             | 6.065        | Positive |
| 50                     |     | 2.57                        | 0.610                             | 4.226        | -       |
| 53                     |     | 2.60                        | 0.947                             | 6.062        | Positive |
| 57                     |     | 2.64                        | 0.595                             | 4.147        | -       |
| 61                     |     | 3.22                        | 0.514                             | 3.690        | -       |
| 62                     |     | 3.23                        | 0.824                             | 5.390        | -       |
| 64                     |     | 3.25                        | 0.794                             | 5.228        | -       |
| 65                     |     | 3.26                        | 0.533                             | 3.796        | -       |
| 68                     | X   | 3.29                        | 0.998**                           | 6.338        | Positive |
| 69                     |     | 3.30                        | 0.531                             | 3.783        | -       |
| 71                     | X   | 3.32                        | 0.001                             | 0.367        | Negative |
| 72                     | X   | 3.33                        | 0.094                             | 1.351        | -       |
| 77                     |     | 3.38                        | 0.776                             | 5.126        | -       |
| 110                    |     | 4.56                        | 0.930                             | 5.967        | -       |
| 114                    | X   | 4.60                        | 0.010                             | 0.556        | Negative |
| 117                    | X   | 5.35                        | 0.253                             | 2.213        | -       |
| 121                    | X   | 5.39                        | 0.001                             | 0.336        | Negative |
| 124                    | X   | 5.42                        | 0.215                             | 2.044        | -       |
| 168                    | X   | 6.55                        | 0.831                             | 5.429        | -       |
| 171                    | X   | 7.35                        | 0.127                             | 1.509        | -       |
| 175                    | X   | 7.39                        | 0.222                             | 2.087        | -       |
| 179                    |     | 7.43                        | 0.574                             | 4.022        | -       |

Model M8 NEB values obtained from subgroup 43. Key position 3.29 is under positive selection (** denotes statistically significant posterior probability for $\omega$>1). Two non-key positions, 2.56 and 2.60, have posterior probability exceeding 90% for positive selection. All positions with posterior probability for $\omega$>1 which exceed 0.5 are represented. Results of $\omega$ from the 10 key positions are also included. Key positions identified in Reference [13,14] are indicated by X. Statistics of the 3 positions under positive selection are represented in bold italics.

doi:10.1371/journal.pone.0027813.t006
be clearly identified as distinct clades from a familial phylogenetic tree. We used the counting scheme of Nei-Gojobori to estimate the average $d_N$ and $d_S$ from pairs of sequences [64]. We investigated subgroups where the maximum average $d_S$ of all pair-wise comparisons within the subgroup did not exceed 1. If the condition of max($d_S$)<1 was not met, then the out group taxa was removed, and the subgroup reduced. There was no a priori scheme to identify subgroups to achieve the max($d_S$) and max($d_S$) conditions. To study the measurement uncertainties due to sample size, we analyzed subgroups having progressively larger numbers of closely related receptors. The subgroups in which it exceeded 1 were indicated by “N” in Tables 1, 2, 3 and 4 and were not included in Figure S4, Figure S6 and Figure S8. We found that max($d_S$)<1 selection resulted in max($d_S$)<3 for forty of forty-five subgroups. Subgroups listed in Tables 1, 2, 3 and 4 and denoted by “S” did not meet max($d_S$)<3. The $d_S$ and $d_S$ obtained after maximum likelihood computation was more conservative compared to that obtained via the Nei-Gojobori counting method (results not shown).

Estimation of $\omega$ at AA positions across 7TMs

PAML version 4.2b [65] was used to model the evolution of the 7TM nucleotide sequences using a state space of possible codons from the genetic code. The program simulated the molecular evolution of the concatenated 7TM fragments independently, for each subgroup. Four independent strategies from PAML were used to estimate $\omega$. Two mathematical models were tested for statistical tenability in each strategy. The constraints and assumptions for estimating $\omega$ were accommodated differently in the models. In the first strategy, model M2a accommodated positions under negative selection via $\omega = \omega_0$ ($\omega_0<1$), a free parameter determined from data, that was common for most 7TM positions. In addition, to represent neutral evolution, a portion of the remaining 7TM positions were constrained to $\omega = 1$. Lastly, with another free parameter, the same model also accommodated representation of positive selection for the remaining fraction of positions ($\omega_2>1$). In contrast, model M1a was a special case of M2a, in which it excluded positive selection. Because $\omega$ for an AA position under near-neutral evolution was also constrained to unity, this was the most conservative of the three strategies. Test 1 compares M1a vs. M2a.

In the second strategy the spectrum of $\omega$ values from MSA positions was represented by a beta function (with two free parameters $p$ and $q$). Model M8 represented the spectrum of $\omega$ across all MSA positions with ten discrete $\omega_i$ categories to represent the beta function (for $\omega_i\leq 1$, $i = 0,1,2...9$). An additional eleventh category $\omega_{10}$ accounted for a small fraction of positions under positive selection. In model M7, there was no provision for such positive selection ($p_{10} = 0$, therefore $\omega_{10}$ was absent). Test 2 compares M7 vs. M8.

In a third strategy, we used Yang and Swanson’s “fixed sites” models A and B [25]. The null model (model A) hypothesized that there was no statistically distinct selection pressure among the MSA positions. We used the simplest alternate model (model B), from the suite of “fixed sites” models, which hypothesized that the average evolutionary selection pressure from cohort of key MSA positions was statistically distinct with respect to the other MSA positions.

In all the three strategies, which we refer to as Tests 1–3 in Table S1, a maximum likelihood ratio test was used to determine the tenable model from competing nested paired models. The goal of both models was to represent the observed evolutionary data – the MSA of nucleotide 7TM sequences and the phylogenetic tree from the relevant subgroup. In each strategy, the maximum...
likelihood of the null model MNull that could fit the data was compared with that obtained from an alternate model MAlt (which had additional free parameters compared to the null model).

In a fourth strategy, which we called Test 4, model M3 was compared to model M0 for all subgroups. The alternative model demonstrated the heterogeneity of \( \phi \) values across the 7TMs and the null model was representative of their common \( \phi \) value. Test 4 is not specific for inferring positive selection and all results are shown in Table S2.

### Chemical class of the natural ligands associated with class A GPCRs

Subgroups were classified into three categories in terms of their natural ligands: 1) small molecules (including biogenic amines, nucleosides and nucleotides), 2) lipids and 3) peptides. If subgroups did not exclusively bind the same chemical class of natural ligand or if they had more than two orphan receptors, then we categorized them as divergent. If subgroups exclusively contained orphan receptors then they were categorized as orphan.

### Computing average \( \phi \) from randomly selected 7TM AA positions

To compare \( \phi \) with randomly selected 7TM positions, two hundred cohorts of AA positions were simulated. The average \( \phi \) from each of the cohort of ten randomly selected 7TM positions was computed – this was denoted as \( \phi_{\text{random7TM}} \). The average of the two hundred independent cohorts was computed from the distribution of \( \phi_{\text{random7TM}} \).

### Computing AA diversity at key positions

Shannon entropy was first used to estimate the diversity in AA composition at key positions across all subgroups. The Shannon entropy at MSA position \( X \), with AA residues \( x \), was defined as

\[
H(X) = \sum p(x) \log_2 p(x).
\]

Here the summation is over all rows \( r \) of the MSA, \( p(x) \) was the probability of having residue \( x \) at position \( X \), and the summation is over all AA residues.

A variety of strategies exist to quantify sequence similarity [66]. We used two independent approaches to estimate the similarity of key AA residues using all subgroups. In the first method, sequence similarity was estimated with the BLOSUM substitution matrix [26]. Consider \( S \) to be the number of concatenated 7TM fragments in a subgroup. The AA similarity (and dissimilarity) among MSA positions of 7TM fragments due to substitutions among the \( S \) different paralogs of the subgroup was determined. We used BLOSUM80 substitution matrix to evaluate sequence similarity among the residues at key positions of the MSA. For a given key position, the average score of the key AA residues substituting with each other within the subgroup, we used the definition of Karlin and Brocchieri [67], given by the equation

\[
C_{\text{Karlin}}(X) = \frac{2}{S(S-1)} \sum \sum \sum \sum M_{x,y},
\]

where \( M_{x,y} \) is the AA at MSA position (or column) \( X \) in the \( i \)th row to AA \( y \) in the \( j \)th row, and \( M_{x,y} \) scores for substitution between AA \( x \) and AA \( y \). This similarity score \( M_{x,y} \), for the defined \( (i,j) \) pairs of AAs in the \( i \)th and \( j \)th sequence fragment, is defined as

\[
M_{x,y} = \frac{m_{x,y}}{\sqrt{m_{x,x}m_{y,y}}},
\]

where \( m_{x,y} \) is the BLOSUM80 [26] matrix element corresponding to substitution from AA \( x \) in the \( i \)th row to AA \( y \) in the \( j \)th row of the alignment (or vice versa). We defined the BLOSUM similarity score for a given key position \( X \) as \( BLO_80_{\text{score}} = C_{\text{Karlin}}(X) \), and the average similarity score of all key positions \( BLO_80_{\text{all}} \) was averaged over the ten key positions.

In another approach, another estimate for dissimilarity was obtained using residues from MSA columns at key positions. To represent a distance measure, the average percentage of accepted mutation using program PROTDIST from PHYLIP software [61] was obtained for all key positions in subgroups. That measure was denoted as \( D_{\text{key}} \). The quantity \( -\log(D_{\text{key}}) \) was computed to

---

**Figure 3.** Graph showing the trends in \( <\phi_{\text{key}}>/<\phi_{\text{random7TM}}> \) vs. \( <\phi> \). Subgroups with pair-wise max\( (d_{xy})<1 \) are represented in these panels. Subgroups 38 and 44 are excluded to avoid bias due to positive selection. **a)** Plot of \( <\phi_{\text{key}}>/<\phi_{\text{random7TM}}> \) vs. \( <\phi> \). **b)** Plot of \( <\phi_{\text{key}}>/<\phi_{\text{random7TM}}> \) vs. \( <\phi_{\text{random7TM}}> \).

doi:10.1371/journal.pone.0027813.g003
compare the attribute with previously computed measures of sequence similarity.

Supporting Information

Figure S1 The average \( \omega \) from key positions (\( \omega_{\text{key}} \)) contrasted with average \( \omega \) from randomly selected 7TM positions (\( \omega_{\text{random7TM}} \)). Results of selection pressure, from PAML’s model M7 vs. M8, for subgroups 1–45, as listed in Tables 1, 2, 3 and 4 of manuscript, are shown. The \( \omega \) values on the Y axis are represented in a linear scale (panel A) and logarithmic scale (panel B – Figure 2 in manuscript). Subgroups from 1–10 (shown in Table 1) are receptors naturally activated by small molecules, 11–19 (shown in Table 2) by lipids and 20–38 (shown in Table 3) by peptides. Subgroups 39–44 (shown in Table 4) are divergent. Subgroup 45 exclusively contains orphan GPCRs. Filled (red colored) triangle represents \( \omega_{\text{key}} \) while open triangle represents the average from random cohorts (from \( \omega_{\text{random7TM}} \) distribution). The error bar represents two standard deviations (2\( \sigma_{\text{random7TM}} \)) or the limits of 95% confidence interval from \( \omega_{\text{random7TM}} \) distribution.

(TIFF)

Figure S2 Graph of \( \omega_{\text{key}} \) vs. \( \omega_{\text{random7TM}} \). Trend from \( \omega_{\text{key}} \) vs. \( \omega_{\text{random7TM}} \) is shown using a logarithmic scale. Graph excludes subgroups labeled as “N” in Tables 1, 2, 3 and 4 and excludes subgroups 38 and 44.

(TIFF)

Figure S3 Shannon entropy (\( H \)) for key positions across GPCR subgroups.

(TIFF)
**Figure S4** Average Shannon entropy vs. average selection pressure for key positions across subgroups. Average scores from Figure S3 are plotted along the Y axis. Average evolutionary selection pressure from Figure 1 is represented using a logarithmic scale on the X axis. Subgroups not labeled “N” from Tables 1, 2, 3 and 4 (having pair-wise \(\text{max}(d_S) < 1\)) are represented here. (TIFF)

**Figure S5** Similarity scores for key positions across GPCR subgroups. Similarity scores \(<\text{BLO}_{80}\text{key}>\) in subgroup MSA defined by Karlin and Brocchieri, as in Reference 67, (described in Materials and Methods) generated using BLOSUM80 matrix. (TIFF)

**Figure S6** Average similarity score \(<\text{BLO}_{80}\text{key}>\) vs. average selection pressure for key positions across subgroups. Average scores from Figure S5 are plotted along the Y axis. Average evolutionary selection pressure from Figure 1 is represented using a logarithmic scale on the X axis. Subgroups not labeled “N” from Tables 1, 2, 3 and 4 (having pair-wise \(\text{max}(d_S) < 1\)) are represented here. (TIFF)

**Figure S7** Inverse protdist distance measure \(<-\log_{10}\text{D}_{\text{key}}>\) for key positions across GPCR subgroups. Plot showing the logarithm of inverse protdist distance \((D)\) at key positions from GPCR subgroups. (TIFF)

**Figure S8** Average inverse protdist distance vs. average selection pressure for key positions across subgroups. The Y-axis represents \(<-\log_{10}\text{D}_{\text{key}}>\) from Figure S7. Average evolutionary selection pressure is represented using a logarithmic scale on the X axis. Subgroups not labeled “N” from Tables 1, 2, 3 and 4 (having pair-wise \(\text{max}(d_S) < 1\)) are represented here. (TIFF)

| Table S1 | Tenable PAML models representing molecular evolution of 7TMs of class A non-olfactory human GPCR subgroups. PAML’s tenable models that represent molecular evolution of their 7TMs are illustrated across GPCR subgroups. Results from two “random sites models” M2a vs. M1a (Test 1), M8 vs. M7 (Test 2) and that from Yang-Swanson “fixed sites” model A vs. model B (Test 3) are presented in columns 5–7. Tenable alternative models are represented “A” and tenable null models labeled “~”. Bold font in column 3 connotes orphan GPCR. Bold and italics in column 5 connotes inference of positive selection. (DOC) |

| Table S2 | Tenable PAML models representing molecular evolution of 7TMs of class A non-olfactory human GPCR subgroups. PAML’s tenable models that represent molecular evolution of their 7TMs are illustrated across GPCR subgroups. Results from “random sites models” M3 vs. M0 (Test 1) are presented. Tenable alternative models are represented “A” and tenable null models labeled “~”. Bold font in column 3 connotes orphan GPCR. Bold and italics in column 5 connotes inference of significant positive selection. (DOC) |

**Acknowledgments**

We acknowledge the use of the IUPHAR database [58,59], NCBI’s online protein and nucleotide database [54,68,69] and ORF resource [55]. We would like to thank Artie Sherman (LBM, NIDDK), Teresa Przytycka (NCBI, NIH), Ivan Ovcharenko (NCBI, NIH) and David Liberles (University of Wyoming, Laramie) for valuable suggestions. We would like to thank Torsten Schoneberg and Eric Vallender for discussions involving results from subgroup of Trace amine receptors. SNF would like to specially thank Joe Bielawski for assistance and discussions with PAML. SNF would like to thank Michael Cummins, Joe Bielawski, Bill Pearson, David Swoford, Joe Feenstra, Mary Kuhner, Michael Miyamoto, Peter Beerrli, Mark Holder and all instructors and TAs from the 2009 Molecular Evolution workshop held at Marine Biological Laboratory, Woods Hole, MA, USA. SNF would also like to thank Tao Tao (NCBI, NIH), Josh Cherry (NCBI, NIH), S. Balaji (formerly from NCBI, NIH) and Adi Stern (Tel Aviv University) for technical assistance. SNF would like to acknowledge the use of computational resources at NIH/NIDDK to accomplish this analysis.

**Author Contributions**

Conceived and designed the experiments: SF SC CC. Analyzed the data: SF SC CC. Wrote the paper: SF SC CC.

**References**

1. Plakidou-Dymock S, Dymock D, Hooley R (1998) A higher plant seven-transmembrane receptor that influences sensitivity to cytokinins. Curr Biol 8: 315–324.
2. Fredriksson R, Lagerstrom MC, Schioth HB (2005) Expansion of the superfamily of G-protein-coupled receptors in chordates. Ann N Y Acad Sci 1040: 89–94.
3. Perez DM (2005) The evolutionarily triumphant G-protein-coupled receptor. Mol Pharmacol 63: 1202–1205.
4. Perez DM (2005) From plants to man: the GPCR “tree of life”. Mol Pharmacol 67: 1383–1384.
5. Schoneberg T, Hofreiter M, Schulz A, Rompler H (2007) Learning from the past: evolution of GPCR functions. Trends Pharmacol Sci 28: 117–121.
6. Takeda S, Kadowaki S, Haga T, Takaesu H, Mitaku S (2002) Identification of G protein-coupled receptor genes from the human genome sequence. FEBS Lett 520: 97–101.
7. Foord SM, Bonner TI, Neubig RR, Rossor EM, Pin JP, et al. (2005) International Union of Pharmacology. XLVI. G protein-coupled receptor list. Pharmacol Rev 57: 279–290.
8. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63: 1256–1272.
9. Archer E, Maigret B, Escrieut C, Pradayrol L, Fourmy D (2003) Rhodopsin crystallization and data collection. Acta Crystallogr D Biol Crystallogr 59: 929–931.
10. Anisimova M, Lunter G, Abecasis GR, Thwaites PJ (2009) A new method for detecting adaptation in rapidly evolving coding sequences using multiple sequence alignments. Mol Phylogenet Evol 53: 300–309.
11. Fredriksson R, Schiöth HB (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes. Mol Pharmacol 67: 1414–1425.
12. Ohno S (1970) Evolution by Gene Duplication. New York: Springer-Verlag.
13. Fatakia SN, Costanzi S, Chow CC (2009) Computing highly correlated positions using mutual information and graph theory for G protein-coupled receptors. PLoS ONE 4: e6681.
14. Fatakia SN, Costanzi S, Chow CC (2011) Comparative genomic analysis using information theory. In: Bhattacharjee MC, Dhar SK, Subramanian S, eds. Recent Advances in Biostatistics: False Discovery, Survival Analysis and other topics. Singapore: World Scientific Press.
15. Hurst LD (2002) The Ka/Ks ratio: diagnostic of the form of sequence evolution. Trends Genet 18: 486.
16. Nekrutenko A, Makova KD, Li WH (2002) The Ka/Ks ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. Genome Res 12: 198–202.
17. Anisimova M, Liberles DA (2007) The quest for natural selection in the age of comparative genomics. Heredity 99: 357–359.
18. Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15: 496–503.
19. Gibson A (2010) Human Evolution. Tracing evolution’s recent fingerprints. Science 329: 740–742.
20. Kosiel C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, et al. (2008) Patterns of positive selection in six Mammalian genomes. PLoS Genet 4: e1000788.
21. Stromberg R, Schrook K, Boett I, Staehert C, Russ A, et al. (2011) Evolution of GPCR: change and continuity. Mol Cell Endocrinol. Ireland: 2010 Elsevier Ireland Ltd. pp 170–178.
22. Choi SS, Lahn BT (2003) Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. Genome Res 13: 2252–2259.
23. Yang S, Liu Y, Liu AA, Cavalli-Sforza LL, Zhao Z, et al. (2005) Adaptive evolution of MRGx2, a human sensory neuron specific gene involved in nociception. Gene 352: 50–58.
24. Ballesteros JA, Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods Neurosci 25: 366–420.
25. Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19: 49–57.
26. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89: 10191–10195.
27. Rompler H, Staubert C, Thor D, Schulz A, Hofreiter M, et al. (2007) G protein-coupled time travel: evolutionary aspects of GPCR research. Mol Interact 7: 17–25.
28. Ault AD, Besch JF (2006) Creation of GPCR-based chemical sensors by directed evolution in yeast. Protein Eng Des Sel 19: 1–8.
29. Rodriguez GF, Yao R, Lichtenberg O, Wensel TG (2010) Evolution-guided discovery and encoding of allosteric pathway specificity determinants in psychoactive hallucinogen receptors. Proc Natl Acad Sci U S A 107: 7787–7792.
30. Staubert C, Bosel I, Bohneck J, Rompler H, Enard W, et al. (2010) Structural and functional evolution of the trace amine-associated receptors TAAR3, TAAR4 and TAAR5 in primates. PLoS One 5: e11133.
31. Vallender EJ, Xie Z, Westmoreland SV, Miller GM, et al. (2010) Functional evolution of the trace amine associated receptors in mammals and the loss of TAAR1 in dogs. BMC Evol Biol 10: 51.
32. Huzurbazar S, Kolesov G, Massey SE, Harris KG, Churbanov A, et al. (2010) Lineage-specific differences in the amino acid substitution process. J Mol Biol 396: 1410–1421.
33. Nei M (1975) Molecular Population Genetics and Evolution. Amsterdam: North-Holland.
34. Nei M (1983) Genetic polymorphism and the role of mutation in evolution; Nei M, Koehn R, eds. Sunderland, MA: Sinauer Associates.
35. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 217–230.
36. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282.
37. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. J Mol Biol 165: 106–112.
38. Schoneberg T, Hermsdorf D, Engenmaier E, Engel K, Liebscher I, et al. (2007) Structural and functional evolution of the P2Y(12)-like receptor group. Purinergic Signal 3: 255–268.
39. Pal C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Mol Biol Evol 18: 2323–2326.
40. Schoneberg T, Hermsdorf D, Engenmaier E, Engel K, Liebscher I, et al. (2007) Structural and functional evolution of the P2Y-like G-protein-coupled receptor. J Biol Chem 272: 35531–35541.
41. Pal C, Papp B, Hurst LD (2003) Does the recombination rate affect the evolution of protein families. Nat Rev Mol Cell Biol England. pp 709–720.
42. Worth CL, Gong S, Blundell TL (2009) Structural and functional restraints in the evolution of protein families and superfamilies. Nat Rev Genet 10: 77–87.
43. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, et al. (2000) The crystal structure of rhodopsin: a G protein-coupled receptor. Science 289: 739–745.
44. Choi SS, Vallender EJ, Lahn BT (2006) Systematically assessing the influence of three-dimensional structural context on the molecular evolution of mammalian proteomes. Mol Biol Evol, United States. pp 2131–2133.
45. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 217–230.
46. Mundy NL, Keedy J (2003) Evolution of a pigment gene, the melanocortin-1 receptor, in primates. Am J Phys Anthropol 121: 67–80.
47. Keller N, Halford S, Foster RG (2009) The evolution of irradiance detection: melanopsin and the non-visual spins. Philos Trans R Soc Lond B Biol Sci 364: 2489–2565.
48. Pal C, Papp B, Hurst LD (2003) The structural evolution of a P2Y-like G-protein-coupled receptor. J Biol Chem 278: 35531–35541.
49. Pal C, Papp B, Hurst LD (2001) Does the recombination rate affect the evolution of protein families. Nat Rev Mol Cell Biol England. pp 709–720.
50. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, et al. (2000) The crystal structure of rhodopsin: a G protein-coupled receptor. Science 289: 739–745.
51. Vallender EJ, Xie Z, Westmoreland SV, Miller GM, et al. (2010) Functional evolution of the trace amine associated receptors in mammals and the loss of TAAR1 in dogs. BMC Evol Biol 10: 51.
52. Worth CL, Gong S, Blundell TL (2009) Structural and functional restraints in the evolution of protein families and superfamilies. Nat Rev Genet 10: 77–87.
53. Nei M, Koehn R, eds. Sunderland, MA: Sinauer Associates.
54. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. J Mol Biol 165: 106–112.
55. Tatusova T, Tatusov R. NCBI Open Reading Frame (ORF) online resource toolkit. Available: http://www.ncbi.nlm.nih.gov/projects/orf/. Bethesda, Maryland, USA.
56. Rosenthal GM, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35: D61–65.
57. UnitProtCoorutum (2009) The Universal Protein Resource (UniProt) 2009. Nucleic Acids Res 37: D169–174.
58. Harmar AJ, Hills RA, Rosser EM, Jones M, Buneman OP, et al. (2009) IUPHAR-db: the IUPHAR database of G protein-coupled receptors and ion channels. Nucleic Acids Res 37: D60-6–685.
59. IUPHAR 7TM receptor database. Available: http://www.iuphar-db.org/ DATABASE/GPCRsForward.
60. Glajcar D, Fredriksson R, Schiöth HB (2007) The G protein-coupled receptor subset of the rat genome. BMC Genomics 8: 338.
61. Feherntein J (1989) PHYLLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.
62. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282.
63. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.
64. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 410–426.
65. Yang Z (2006) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586–1591.
66. Valtart W (2002) Scoring residue conservation. Proteins 48: 227–241.
67. Karlin S, Brocchiari L, B. Evolutionary conservation of RecA genes in relation to protein structure and function. J Bacteriol 178: 1881–1894.
68. NCBI gene database. Available: http://www.ncbi.nlm.nih.gov/sites/entrez?db = gene. Bethesda, Maryland, USA.
69. NCBI protein database. Available: http://www.ncbi.nlm.nih.gov/protein/. Bethesda, Maryland, USA.
70. Li J, Edwards FC, Burghammer M, Villa C, Schertler GF (2004) Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol 343: 1409–1430.