



**HAL**  
open science

## Evolution of GPCR: change and continuity

Rainer Strotmann, Kristin Schröck, Iris Bösel, Claudia Stäubert, Andreas Russ, Torsten Schöneberg

► **To cite this version:**

Rainer Strotmann, Kristin Schröck, Iris Bösel, Claudia Stäubert, Andreas Russ, et al.. Evolution of GPCR: change and continuity. *Molecular and Cellular Endocrinology*, 2010, 331 (2), pp.170. 10.1016/j.mce.2010.07.012 . hal-00654483

**HAL Id: hal-00654483**

**<https://hal.science/hal-00654483>**

Submitted on 22 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Title: Evolution of GPCR: change and continuity

Authors: Rainer Strotmann, Kristin Schröck, Iris Bösel, Claudia Stäubert, Andreas Russ, Torsten Schöneberg

PII: S0303-7207(10)00373-4  
DOI: doi:10.1016/j.mce.2010.07.012  
Reference: MCE 7599

To appear in: *Molecular and Cellular Endocrinology*

Received date: 9-12-2009  
Revised date: 28-6-2010  
Accepted date: 13-7-2010

Please cite this article as: Strotmann, R., Schröck, K., Bösel, I., Stäubert, C., Russ, A., Schöneberg, T., Evolution of GPCR: change and continuity, *Molecular and Cellular Endocrinology* (2010), doi:10.1016/j.mce.2010.07.012

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



## Evolution of GPCR: change and continuity

Rainer Strotmann<sup>1</sup>, Kristin Schröck<sup>1</sup>, Iris Bösel<sup>1</sup>, Claudia Stäubert<sup>1</sup>, Andreas Russ<sup>2</sup> and Torsten Schöneberg<sup>1</sup>

<sup>1</sup>Institute of Biochemistry, Medical Faculty, University of Leipzig, Germany

<sup>2</sup>Genetics Unit, Department of Biochemistry, University of Oxford, UK

### *Abstract*

Once introduced into the very early eukaryotic blueprint, seven-transmembrane receptors soon became the central and versatile components of the evolutionary highly successful G protein-coupled transmembrane signaling mechanism. In contrast to all other components of this signal transduction pathway, G protein-coupled receptors (GPCR) evolved in various structural families, eventually comprising hundreds of members in vertebrate genomes. Their functional diversity is in contrast to the conserved transmembrane core and the invariant set of intracellular signaling mechanisms, and it may be the interplay of these properties that is the key to the evolutionary success of GPCR. The GPCR repertoires retrieved from extant vertebrate genomes are the recent endpoints of this long evolutionary process. But the shaping of the fine structure and the repertoire of GPCR is still ongoing, and signatures of recent selection acting on GPCR genes can be made visible by modern population genetic methods. The very dynamic evolution of GPCR can be analyzed from different perspectives: at the levels of sequence comparisons between species from different families, orders and classes, and at the level of populations within a species. Here, we summarize the main conclusions from studies at these different levels with a specific focus on the more recent evolutionary dynamics of GPCR.

### *Introduction*

G protein-coupled receptor (GPCR) genes constitute a substantial fraction of genes in mammalian genomes [1,2]. They are involved in both, sensory functions like photo- or olfactory sensation and regulation of homeostasis including endocrine control of organ function or modulation of neuronal functions. Their abundance in most eukaryotic species raises the question which properties may account for the evolutionary success of this gene family. Phylogenetic analyses demonstrate that the seven-transmembrane (7TM) core structure is highly conserved during GPCR evolution, a feature that can be detected by quantitative bioinformatics as strong purifying (negative) selection. Indeed, the 7TM bundle is used as a structural hallmark for the identification of GPCR in genomic sequences. On the other hand, GPCR bind and respond to ligands highly diverse in size and chemical and physical properties. Long evolutionary processes have shaped and optimized ligand binding and intracellular signaling. Hence, the comparison of orthologous sequences from multiple species has been used to identify functionally relevant residues or motifs and elucidate the molecular basis of high-affinity ligand binding and signal transduction mechanisms [3].

However, our knowledge on GPCR evolution is based on sequences from extant species. As most species that ever lived on earth are extinct, the intermediates of the evolutionary process are no more accessible and we can only infer the structural properties of ancestral sequences. With the current acceleration in high throughput sequencing methods, an increasing number of fully sequenced genomes become available. This greatly enhances the resolution of phylogenetic research. But what information can we extract from this data for the GPCR field?

Firstly, GPCR are evolutionary old. Specific GPCR signatures can be found in all eukaryotic species [4]. This implies that strong negative selection has been continuously acting on

members of this gene family for hundreds of millions of years. Secondly, GPCR have evolved in five structurally distinct gene families since early metazoan evolution [2,5]. The phylogenetic analysis reveals that the number of GPCR in these families differs strongly between classes of species. The available genomic data allows both the quantification and analysis of diversity of the species-specific GPCR repertoires. There are a number of publications on this subject [1,6-10], and in the following section, only a brief outline will be given.

## *GPCR evolution*

### *The origin of the 7TM core*

Signaling through 7TM receptors is a feature of eukaryotic organisms but proteins with a 7TM topology can already be identified in prokaryote genomes. This includes light-sensitive proteo-, bacterio- and halorhodopsins that are involved in non-photosynthetic energy harvesting in archaea and bacteria [11,12]. In some halobacteria, sensory rhodopsins control phototaxis [13]. In contrast to the invariant G protein-dependent intracellular signaling pathway in eukaryotes, retinal photoactivation in prokaryotes results in the formation of transmembrane proton or chloride gradients that drive membrane-bound ATPases, or in interaction with kinases that control flagellar activity.

Although structurally similar sensory rhodopsins are also found in eukaryotes [14,15], their phylogenetic relation to GPCR remains unclear [16]. As the divergence between prokaryotes and eukaryotes has taken place roughly 1.2 billion years ago, the sequence similarity between the 7TM domains is low and a common origin can not be inferred from sequence data.

The answer to how the 7TM core of eukaryotic GPCR has evolved still needs to be found. Interestingly, the highest sequence similarity between the transmembrane helices of bacteriorhodopsin and mammalian GPCR is found between non-homologous helices. This finding has been explained by some authors with an evolutionary mechanism that involves exon shuffling [17]. An alternative hypothesis proposes gene duplication of an ancestral three-transmembrane module that gave rise to both helices 1 through 3 and 5 through 7 [18]. However, it may remain impossible to reconstruct the primordial 7TM prototype. First, both prokaryotic and eukaryotic 7TM proteins have evolved independently for hundreds of millions of years with strongly different generation times, environmental selection, susceptibilities to mutations and modes of genetic rearrangement [18]. Secondly, at least for type 1 microbial rhodopsins, there is good evidence for lateral gene transfer events across great evolutionary distances and even between prokaryotes and eukaryotes that further complicates the analysis of phylogenetic relations [19,20].

As introns in protein-coding regions are uncommon in prokaryotes, extant gene structure analyses can only guide eukaryotic phylogenetic analysis. Indeed, the exon composition of the 7TM core varies strongly both between receptor families and species. In the rhodopsin family, introns within the transmembrane domain are found in only 35.5 % of the human genes whereas 52.5 % are intronless throughout the protein coding sequence [21]. This is in contrast to the invertebrate rhodopsin-family GPCR where the median intron numbers are 2 and 5 for *C. elegans* and *D. melanogaster*, respectively [22]. Phylogenetic analyses suggest that the lower intron number in mammalian species is a consequence of rapid gene expansion through RNA intermediates [21]. However, the intron patterns observed in the ancestral receptor groups do not converge towards a common ancestral gene structure.

A different situation is found in secretin-family GPCR. Here, introns are frequent and specific intron patterns are highly conserved in position and phase between invertebrate and vertebrate species [23]. A high number on introns within the 7TM core is also found in distinct branches of the metabotropic glutamate-receptor and frizzled families, in which introns are otherwise absent or sparse. Notably, in the GABA<sub>B</sub> receptor group, the 7TM

domain contains six introns that are conserved between invertebrate and mammalian species. The same is true for the smoothed receptors from the frizzled receptor family. Despite the strong phylogenetic conservation within these groups, there is no overlap in the gene structure between the GPCR families that would allow the identification of a common ancestral 7TM domain in eukaryote evolution.

### *Evolution of the five GPCR families*

Members of the GPCR superfamily interact with a broad spectrum of extracellular ligands and intracellular signaling molecules. Correspondingly, both the extracellular and intracellular portions of the receptor proteins differ greatly in length and sequence.

Phylogenetic classifications thus rely not only on well-defined functional domains within the N- and C-termini but also on the membrane spanning region that is relatively conserved in all GPCR. Using these criteria, most mammalian GPCR can be assigned to one of the five basic GPCR families: Glutamate (G), Rhodopsin (R), Adhesion (A), Frizzled/Taste2 (F) and Secretin (S) [2].

Genomic data and functional evidence indicate that members of the **glutamate-receptor-like family**, or class C receptors are present in the parazoan *Geodia cydonium* that marks early metazoan evolution and even in the slime mold *Dictyostelium discoideum* whose phylogenetic origin predates metazoan and fungal evolution about 600 million years ago. The glutamate receptor group is thus among the phylogenetically oldest GPCR families.

The distinctive N-terminal Venus fly trap module of the members of this group that is involved in ligand binding has some sequence similarity with the bacterial periplasmic amino acid-binding proteins [24]. Glutamate-dependent activation, however, is absent in *D. discoideum* and occurs only at millimolar concentrations in *G. cydonium*. This suggests that glutamate signaling through this receptor family has arisen early in metazoan evolution [25]. While in the ancestral forms of class C receptors, amino acid (glutamate, GABA) binding was predominant, the glutamate receptor family expanded greatly during vertebrate evolution, giving rise to pheromone-, taste-, calcium-sensing- and fish olfactory receptors.

Conformational changes induced by ligand binding are propagated from the Venus fly trap module via cysteine-rich domains in the N terminus to the 7TM domain. In metabotropic glutamate receptors, the cysteine-rich domains contain nine critical cysteine residues, eight of which are linked by disulfide bridges [26]. Within the transmembrane domain of glutamate-receptor-like GPCR, a number of conserved amino acids are also highly conserved in rhodopsin-like GPCR indicating a relatively close relation between both families. This includes basic amino acids within the third intracellular loop that are involved in receptor activation and a motif in TM7 that is related to the N/DP(X)<sub>n</sub>Y motif in rhodopsin-like GPCR.

The typical signature of members of the **Adhesion GPCR family**, including long N termini with multiple O- and S-glycosylation sites, EGF- and GPCR proteolytic (GPS) domains [27], can be found in most metazoan species. Interestingly, the unicellular choanoflagellate *Monosiga brevicollis* that is regarded as the closest relative of metazoans expresses GPCR with the characteristic GPS domain found in metazoan adhesion GPCR [28,29]. Sequences distantly related to adhesion GPCR are also found in fungal and plant genomes [30,31]. Adhesion GPCR, like glutamate-receptor-like GPCR, may thus have evolved before metazoan evolution. However, experimental evidence for a role in classical G protein-coupled signal transduction is lacking so far for most members although G-protein binding has been shown for the lectomedin receptor and GPR56 [32,33].

In mammals, adhesion GPCR show strong expansion with 31 and 33 members in mouse and humans, respectively [34]. Based on sequence features within the 7TM core domain and domain composition of the N terminus, these receptors can be further grouped into 8 classes.

Interestingly, the number of receptors in the respective classes varies significantly between species with half of the mammalian adhesion GPCR genes being clustered in the genome [35-37]. Both findings point to rapid gene duplication within this family.

In spite of the structural variability, two classes of adhesion GPCR seem to be highly conserved in all bilateral animals and may play a central role in tissue development. This includes the cadherin-like family that comprises the vertebrate CELSR (Cadherin, EGF-like, LAG-like seven-pass receptor) and *Drosophila* flamingo/starry night receptors [38-40]. Members of this family are involved in the control of cell and tissue polarity and interact with wnt/frizzled signaling [41,42].

The second highly conserved class comprises members of the lectin-like (latrophilin) family named after an unusual domain related to rhamnose-binding lectins. Vertebrate in contrast to invertebrate members of this family contain an additional olfactomedin-like domain that has been introduced by a domain shuffling process during vertebrate evolution [6]. Latrotoxin, the neurotoxin of the Black Widow spider, was shown to bind to these receptors but endogenous ligands have not been identified yet. Two of the three mammalian members of this family are specifically expressed in brain and were initially suggested to modulate synaptic exocytosis [43] but may also regulate developmental functions [44].

Phylogenetic analyses of the transmembrane core show a low degree of overall amino acid identity between the paralogs of the adhesion GPCR family. This has been interpreted by some authors as a sign of lacking functional constraint since the 7TM domain is not involved in ligand binding in this family [34]. There are however some highly conserved residues, in particular in TM3, a region that has been shown to be essential for transmembrane signaling in other GPCR families [45].

The **Secretin GPCR family** has been shown to form a phylogenetic group distinct from the adhesion GPCR although both share a similar 7TM core and a conserved N-terminal hormone binding (HRM) domain that is implicated in peptide hormone binding. Secretin GPCR are likely to be descendants of group V adhesion receptors [9] and are found in all bilateria but not in evolutionary older metazoa. Mammalian secretin-family receptors have a complex genomic structure with several introns and are subject to alternative splicing [46]. With respect to their genomic organization, all deuterostome secretin receptor-family orthologs cluster in two families. Phylogenetic comparison to the protostome genes from arthropods and nematodes suggests that the common ancestor was like the calcitonin/CGRP receptors found in vertebrates [23]. All secretin-like GPCR contain a conserved region within the N terminus that plays a pivotal role in ligand binding. This domain includes disulfide bridges between six conserved cysteine residues as well as three clusters of hydrophobic side chains including two conserved tryptophan residues that define its tertiary structure [47]. Between the first and second extracellular loops, two conserved cysteine residues form an additional disulfide bond, a feature that is also conserved in rhodopsin-like GPCR. In contrast to rhodopsin-like GPCR, however, many of which appear to rely on internal hydrophobic sequences for plasma membrane targeting, most secretin-like GPCR possess an N-terminal signal peptide [48].

Members of the **Frizzled-like receptor family** are identified in sponges [49,50], cnidaria [51-53] and placozoa (*Trichoplax adherens*) [54] and have been regarded as an innovation of multicellular metazoan evolution [53]. There is no species specific expansion in number within this family. Among all GPCR, frizzled receptors are the most highly conserved, which is in agreement with their crucial role in eumetazoan ontogenesis including regulation of cell proliferation, cell polarity, gastrulation and tissue formation.

In contrast to this view, the genome of the slime mold *Dictyostelium* that forms a multicellular organism only at certain stages of its lifecycle contains genes that are related to frizzled receptors. Some of these contain N-terminal cysteine-rich and C-terminal KXXXXW domains that mediate Wnt ligand binding and  $\beta$ -catenin-dependent signaling, respectively, in metazoans [55]. Although no functional data is available for the *Dictyostelium* orthologs,

this implies that the phylogenetic origin of the frizzled receptor family precedes metazoan evolution.

The **Rhodopsin-like GPCR family** forms the largest GPCR subfamily in vertebrates, and receptors from this family bind ligands as chemically diverse as small molecular amines, purines, lipids, peptides, large glycoproteins or sense light through a retinal moiety. With the exception of the glycoprotein hormone receptors, ligand binding is thought to occur in the extracellular portion of the transmembrane helix bundle and does not involve the N terminus that is generally shorter than in the other GPCR families.

Rhodopsin-like GPCR feature a conserved TM core with typical E/DRY and N/DP(X)nY motifs on the border of TM3 and ICL2 and TM7, respectively, that are present in almost all members of this family (for more details on conserved structural determinants see [56]). GPCR with this signature are identified in all bilateria but also in cnidaria which implies that rhodopsin-like GPCR evolved before the protostome-deuterostome split, 570-700 Myr ago. The higher number of GPCR in vertebrate compared to invertebrate genomes is mainly caused by a continuous expansion of the rhodopsin-like group during vertebrate evolution [1]. Within rhodopsin-like receptors, a phylogenetically old subfamily can be identified that according to its prominent feature is referred to as leucine-rich repeat-containing GPCR. Members of this group can already be identified in molluscs, nematodes and cnidaria [57-59] and include glycoprotein and relaxin receptors. They share a long N terminus that contains leucine-rich repeats flanked by cysteine-rich sequences [60]. This ligand-binding ectodomain has evolved independently from the 7TM core and its folding properties suggest that it was linked to the 7TM core through genomic rearrangement.

### *GPCR repertoires*

A striking feature of GPCR evolution is their highly dynamic repertoire in eukaryotic organisms. Invertebrate genomes contain the basic ancestral repertoire of vertebrate GPCR genes, and the origin of the five GPCR subfamilies dates back to early metazoan evolution. However, in vertebrate evolution there is a strong expansion in the number and diversity of receptors within all families and particularly in the rhodopsin family [1]. There are notable exceptions like the massive chemosensory receptor inventories that disproportionately contribute to the *C. elegans* [1] and *Anopheles* [61] GPCR portfolios.

Expansion of genes can occur through different genetic mechanisms. Interestingly, the prevalent mechanism in GPCR expansion is gene duplication followed by independent evolution of the copies. This is in contrast to other gene families, like the immune globulins, where diversity is mainly created by genomic rearrangement of hypervariable elements or to other genes where extensive alternative splicing generates proteins with different properties. There is evidence that in early vertebrate evolution, two independent rounds of genome duplication supplied the genomic material for GPCR expansion [62]. These events occurred during chordate evolution after the split of the urochordate and cephalochordate lineages but before the radiation of extant jawed vertebrates [63]. An additional whole genome duplication may have occurred in the teleost lineage after its split from the tetrapod lineage, from which only a subset has been retained in modern teleost genomes [64]. *Coelacanth*, that is considered a close relative to tetrapods, however shows no evidence for whole genome duplication [65]. It is interesting that after whole genome duplication events, non-chemosensory GPCR genes were retained with a higher probability than expected [66]. This may indicate that GPCR signaling *as such* is positively selected for, a finding that corresponds with the general expansion of GPCR in more complex species.

Alternative duplication events may affect single genes or whole chromosomes. Indeed, conserved synteny indicates that chromosomal duplication generated a set of paralogous genes on chromosomes 4 and 5 including the dopamine and adrenergic receptor paralogs, DRD5 and DRD1 and ADRA2C and ADRA1, respectively [67].

A number of GPCR subfamilies that were subject to independent rapid yet species-specific expansion, including olfactory [68-70], chemokine [71], aminergic [72], trace amine-associated (TAAR) [73,74], vomeronasal [75] and nucleotide receptor-like receptors [76], occur in clusters within the vertebral genomes and are often arranged in a tandem fashion. This implies a duplication mechanism that is confined to a genomic region. Mechanisms of this kind are extensively studied in prokaryotes and may involve imperfect recombination events during genome replication with or without circular DNA intermediates [77]. There is evidence that similar mechanisms may underlie gene amplification in eukaryotes [78].

An interesting theory on the specific expansion of olfactory, chemosensory, arthropod gustatory and pheromone receptors was proposed by Fredriksson et al. [1]. These GPCR share a compact structure with short N termini and small molecule ligands. This may reduce the structural constraints for ligand binding and enhance the evolutionary survival of duplicated genes.

On the phylogenetic scale, there are also lineage-specific episodes of loss of receptor diversity and pseudogenization. Gene duplicates are subject to opposing evolutionary forces. Duplicates may mutate and acquire new functionality that drives positive selection while the ancestral paralog maintains the ancestral function. On the other hand, disadvantageous mutants will be eliminated from the population and neutral evolution may lead to the elimination of gene copies that contain non-advantageous mutations. Release of selective pressure results in the acquisition of further mutations which increases the ratio of non-synonymous to synonymous substitution rates ( $K_a/K_s$ ) in the gene duplicate [79]. Eventually, products of this process can be identified in the genome as pseudogenes in which the ancestral gene can still be recognized but nonsense or frame shift mutations are unambiguous signs of functional inactivation. In the case of TAAR, phylogenetic estimates suggest that accumulation of such obvious features takes roughly 7-10 Myr of neutral evolution [80]. However, missense mutations that impair protein function may render a gene functionally inactive although there is none of the critical pseudogene features present. This has been described for a number of GPCR including chemoattractant and olfactory receptors [81-83]. Unless tested in functional assays, these pseudogenes thus escape detection. Consequently, the number of pseudogenes may be strongly underestimated by genomic analyses alone.

In humans, pseudogenization is particularly frequent in odorant GPCR (OR) with 63 % of the roughly 900 OR showing obvious structural features of inactivation [84]. In the mouse genome, only 20 % of the OR genes are pseudogenes [85]. A similar degree of pseudogenization has been observed for human bitter taste and pheromone receptors [86,87]. It has been speculated that the evolution of trichromatic color vision has relaxed the functional constraints for these receptor groups in hominoids and other Old World monkeys [88-90].

What are the evolutionary advantages of an increase in GPCR number? Comparing invertebrates and vertebrates, one can observe a correlation between the repertoire of GPCR families and complexity of the organism [10]. However, most GPCR do not play a primary vital role in the organism. More than 50 % of knockout mice of individual GPCR display no or a moderate phenotype unless challenged with extreme conditions [91,92]. A larger GPCR repertoire provides the living organism with more sensory information towards the environment or allows for more complex homeostatic regulation. This may permit better behavioral adaptation to environmental conditions. A larger number of GPCR genes also supplies more genetic material that may be subject to evolutionary changes and enhance adaptation on the evolutionary level.

In summary, phylogenetic analyses are important tools in GPCR research that have given us information about the origin of GPCR, the evolution of the five subgroups and structure-function relations in ligand binding and signaling. However, it is difficult to reconstruct the



evolutionary forces that have shaped the receptor repertoires from phylogenetic comparisons alone.

### *Signatures of selection*

Extant species define the current endpoints of long lineage-specific evolution that involved genetic drift as well as selection. Indeed, many evolutionary old GPCR genes including the rhodopsins share strong conservation between distantly related species and have low  $K_a/K_s$  ratios throughout their coding regions particularly in the highly conserved 7TM region [86,93,94]. This finding is indicative of continuous negative selection. However, it is not always possible to identify true orthologs from distantly related species like, for example, vertebrates and invertebrates. During their independent evolution, different selective pressures have acted on orthologs, producing species-specific receptor repertoires, and ligand-receptor co-evolution may have shaped the specific receptor response to a given ligand.

Likewise, the positive selection mechanisms that led to fixation of paralogous genes or pseudogenization can not be reproduced from phylogenetic sequence comparisons. Some studies have tried to infer causalities, for example the loss of olfactory GPCR [95] in response to the advent of trichromatic vision in primates [96]. Large scale degeneration of olfactory GPCR genes is found in cetaceans (dolphins, whales), which have secondarily adapted to a marine habitat and have lost or greatly reduced their sense of smell acquired in terrestrial environments [97].

Taken together, using phylogenetic data, it is difficult to prove how variation of a gene has generated an evolutionary advantage under certain environmental circumstances. Therefore, the identification of specific selection mechanisms remains the most challenging task in evolution research.

### *The population genetics view*

A complementary approach to phylogenetic sequence comparisons is the analysis of genetic variation within a population. Population genetics models predict that rapid selection leaves footprints in closely linked genomic regions (selective sweeps) that can be detected in single nucleotide polymorphism (SNP) data by bioinformatics methods [98]. Recently, large-scale SNP datasets from selected human population samples [99] and other species [100] have been made available in the International Haplotype Map (HapMap), Perlegen Sciences and other projects. They are used to scan the human genome for signatures of recent positive selection [101]. Commonly analyzed parameters include the reduction in genetic diversity, frequency of derived alleles, allele frequency differences between populations and haplotype length (for further reading see [102]). Such analyses allow the identification of signatures of selection within one species even before selected genes become fixed in the population. Furthermore, SNP analysis does not discriminate between coding and non-coding regions and can thus identify natural selection irrespective of the genomic area. However, a challenging problem of this approach remains the clear distinction between positive selection and neutral effects like genetic drift, bottleneck-, founder- and other demographic phenomena [81].

The meta-analysis of genome-wide SNP studies in humans reveals a number of genomic regions with strong signatures of positive selection that contain GPCR genes (Table 1). This includes chemosensory receptors, receptors that are involved in immune defense and, as expected from phylogenetic data, receptors of the adhesion receptor family (BAI3, CELSR1, GPR64/HE6). However, owing to methodological reasons, this approach has a limited genomic resolution and can not identify the action of selection on separate genes.

Interspecies and population genetic analyses were applied also to other species in order to extract information on adaptive changes in GPCR. In *Drosophila* for example, the *methusalah*

(*mth*) gene codes for a GPCR that has been associated with regulation of stress response and lifespan [103]. A population genetics study in different *Drosophila* species [104] revealed that there is both unusually high divergence in the intra- and extracellular loop domains of the receptor protein between the closely related species, and signatures of temporally and spatially variable selection on the *mth* locus within different populations of *D. melanogaster*. Both findings strongly suggest recent adaptive evolution of the *mth* GPCR gene.

SNP datasets can not only be used for the identification of genomic regions that have been subject to recent selection but also to correlate pheno- and genotypical features in a population sample. Several large-scale genome-wide association (GWA) studies in human cohorts, phenotyped for diseases or disease predispositions, and in mice have been recently performed (further readings [105,106]). They often revealed known, and sometimes formerly unrecognized genes/loci that are parts of pathways involved in complex phenotypic traits. Not unexpectedly, GPCR are frequently found in such hypothesis-free analyses. For example, variations in the thyrotropin-releasing hormone receptor- and GPR133 genes have been shown to be associated with low lean body mass [107] and body height [108].

The analysis of genetic variation in population samples has doubtlessly proven its power to point at genes that are potentially involved in complex functions. This enhances the methodical repertoire in a way that is more physiological than experimental genetic methods in which specific gene inactivation in animals represents an extreme functional endpoint.

### *Selection acting on GPCR genes*

Genome-wide approaches usually identify genomic regions/loci where specific signatures suggest the action of selection, but not specific genes. To detect evolutionary selection on individual genes, other methods must be employed. A number of studies demonstrate increases in frequency of distinct non-synonymous polymorphisms that are located within the coding region of odorant and taste receptors in human and other vertebrate populations [87,109-111]. In many cases, this was implicated in altered ligand binding or signal transduction properties in the derived allele. For example, the bitter-taste receptor TAS2R16 [111] like other bitter-taste receptors [110] was found to display an excess of evolutionary derived alleles in the human population at the non-synonymous site Lys<sup>172</sup>Asn. Functional studies demonstrated that the derived allele has a higher sensitivity toward harmful cyanogenic glycosides than the ancestral mammalian variant [111]. The improved recognition of natural toxins may have conferred an important advantage at the early stage of human evolution and thus driven the increase in frequency of the Asn<sup>172</sup> allele in the human population.

A well documented example for selection is the melanocortin type 1 receptor (MC1R) gene. MC1R controls pigmentation of skin and hair [112]. In African populations, strong purifying selection is found in the *Mc1r* gene whereas there is a substantial increase in genetic diversity in Europeans and Asians. Many of the non-African variants are associated with reduced intracellular signaling which leads to lighter skin and hair color [113,114]. This is commonly interpreted as a consequence of relaxed constraint considering that protection from sunlight is less relevant in non-equatorial regions [113,115]. It is unclear whether positive selection of dysfunctional alleles may have played an additional role in MC1R evolution. Alleles associated with reduced MC1R activity were also found in pleistocene species including Neanderthals [116] and mammoths [117] in which reduced pigmentation may have promoted cutaneous vitamin D synthesis under extreme climate conditions during ice ages. The opposite situation was found in rock pocket mice that are generally light-colored and live on light-colored rocks. However, populations of dark (melanic) mice are found on dark lava substrates. Association studies identified four non-synonymous sites within the *Mc1r* gene that are present at high frequency in the melanic population [118].

Another example refers to the V2 vasopressin receptor (V2R) which regulates renal water reabsorption. There is evidence that structural changes in the marsupial V2R resulted in a

gain of V2R function and thus may be advantageous in maintaining water and electrolyte homeostasis for marsupials living in arid habitats [119].

### ***Positive selection on the null-allele***

The clearest lines of evidences for positive selection come from loss-of-function gene variants. Evolutionary advantage through loss-of-function in MC1R has been clearly shown for beach mice in Florida [120,121] and a lizard species in the Chihuahuan Desert [122,123]. Both species have specifically adapted to their sand dune habitats by a lighter coat color than their wild type counterparts. As shown in Table 1, genomic regions with clear signatures of selection contain a number of GPCR pseudogenes. Thus, positive selection on null alleles may play a more important role in adaptation than previously appreciated.

As another example, the FY\*O allele at the Duffy locus is at or near fixation in sub-Saharan Africa but rare in other parts of the world [124]. The Duffy antigen which is a chemokine receptor acts as a co-receptor for the cell entry of *Plasmodium vivax*. Mutation studies have shown that inactivation of Duffy leads to resistance to infections by *P. vivax* [125]. Selection on the null-allele is probably involved in the evolution of the chemoattractant-like receptor GPR33 [81,83]. In the chemoattractant-like receptor GPR33, that appeared in the mammalian genome more than 125-190 Myr ago, pseudogenization was found in humans, which occurred less than 1 Myr ago. Interestingly, independent inactivation events could be identified in hominoids and some rodent species [81] in a similar timeframe. As simultaneous pseudogenization in different species due to neutral drift is very unlikely, a common selection process like a rodent-hominoid-specific pathogen was speculated [81]. A similar situation was found in TAAR that possibly function as chemosensory receptors in olfactory epithelia [74,126]. Two paralogs of this family, TAAR3 and TAAR4 independently became pseudogenized in great apes and New World monkeys on a comparable time scale on different continents (Stäubert/Schöneberg unpublished data). Given their coincidence, a common exogenous trigger is a likely cause for these unconnected events.

### ***Conclusion***

The recent advances in genotyping and next generation sequencing methods have greatly accelerated the availability of genetic data from both a growing number of species as well as individuals from different populations within one species. When sequence information from different species first became available, evolutionary GPCR research focused on the phylogeny of 7TM proteins and the origin of GPCR. From this, we have learned which GPCR families are evolutionary old and which receptor groups underwent substantial expansion and pseudogenization during evolution. One interesting conclusion is that, while some GPCR appear to be essential for eukaryotic life, the distinctive property of GPCR genes to readily adapt to new sensory functions may represent an evolutionary advantage *per se*.

However, one general drawback of the phylogenetic approach is the lack of information, what evolutionary forces and environmental conditions have shaped the species-specific GPCR repertoires. Population genetics data provides a complementary concept. Although the evolutionary distance between population groups of one species is smaller and the extent of environmental adaptation is lower and thus more difficult to detect, selection processes can be quantitatively observed by bioinformatics methods.

Genome-wide association studies have supplied candidate genes for GPCR-associated phenotypes and sequence analyses have identified signatures of selection in distinct genomic regions. Accordingly, these methods are suitable tools for the identification of previously unrecognized roles of GPCR and for the generation of hypotheses on how adaptation occurs on the gene level. However, the holy grail of GPCR genetics remains the demonstration of a causal link between genetic variation and altered receptor function. To this end, functional experiments are indispensable and studies that clearly demonstrate species adaptation to

environmental conditions on a molecular level are still sparse. In future studies, full-genome sequences instead of genotyping markers like SNP will greatly enhance the resolution of population genetics studies and identify polymorphisms in GPCR genes that have been previously undetected or that represent rare alleles.

## References

- [1] Fredriksson, R. and Schioth, H.B. (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol Pharmacol* 67, 1414-25.
- [2] Schioth, H.B. and Fredriksson, R. (2005) The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen Comp Endocrinol* 142, 94-101.
- [3] Rompler, H., Staubert, C., Thor, D., Schulz, A., Hofreiter, M. and Schoneberg, T. (2007) G protein-coupled time travel: evolutionary aspects of GPCR research. *Mol Interv* 7, 17-25.
- [4] Schoneberg, T., Hofreiter, M., Schulz, A. and Rompler, H. (2007) Learning from the past: evolution of GPCR functions. *Trends Pharmacol Sci* 28, 117-21.
- [5] Josefsson, L.G. (1999) Evidence for kinship between diverse G-protein coupled receptors. *Gene* 239, 333-40.
- [6] Bjarnadottir, T.K., Fredriksson, R. and Schioth, H.B. (2007) The adhesion GPCRs: a unique family of G protein-coupled receptors with important roles in both central and peripheral tissues. *Cell Mol Life Sci* 64, 2104-19.
- [7] Schioth, H.B., Nordstrom, K.J. and Fredriksson, R. (2007) Mining the gene repertoire and ESTs for G protein-coupled receptors with evolutionary perspective. *Acta Physiol (Oxf)* 190, 21-31.
- [8] Nei, M., Niimura, Y. and Nozawa, M. (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9, 951-63.
- [9] Nordstrom, K.J., Lagerstrom, M.C., Waller, L.M., Fredriksson, R. and Schioth, H.B. (2009) The Secretin GPCRs descended from the family of Adhesion GPCRs. *Mol Biol Evol* 26, 71-84.
- [10] Fredriksson, R., Lagerstrom, M.C. and Schioth, H.B. (2005) Expansion of the superfamily of G-protein-coupled receptors in chordates. *Ann N Y Acad Sci* 1040, 89-94.
- [11] Oesterhelt, D. (1998) The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr Opin Struct Biol* 8, 489-500.
- [12] Fuhrman, J.A., Schwalbach, M.S. and Stingl, U. (2008) Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* 6, 488-94.
- [13] Spudich, J.L. and Luecke, H. (2002) Sensory rhodopsin II: functional insights from structure. *Curr Opin Struct Biol* 12, 540-6.
- [14] Sineshchekov, O.A., Jung, K.H. and Spudich, J.L. (2002) Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* 99, 8689-94.
- [15] Waschuk, S.A., Bezerra, A.G., Jr., Shi, L. and Brown, L.S. (2005) *Leptosphaeria* rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote. *Proc Natl Acad Sci U S A* 102, 6879-83.
- [16] Soppa, J. (1994) Two hypotheses--one answer. Sequence comparison does not support an evolutionary link between halobacterial retinal proteins including bacteriorhodopsin and eukaryotic G-protein-coupled receptors. *FEBS Lett* 342, 7-11.
- [17] Pardo, L., Ballesteros, J.A., Osman, R. and Weinstein, H. (1992) On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors. *Proc Natl Acad Sci U S A* 89, 4009-12.

- [18] Taylor, E.W. and Agarwal, A. (1993) Sequence homology between bacteriorhodopsin and G-protein coupled receptors: exon shuffling or evolution by duplication? *FEBS Lett* 325, 161-6.
- [19] Sharma, A.K., Spudich, J.L. and Doolittle, W.F. (2006) Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* 14, 463-9.
- [20] Ruiz-Gonzalez, M.X. and Marin, I. (2004) New insights into the evolutionary history of type 1 rhodopsins. *J Mol Evol* 58, 348-58.
- [21] Fridmanis, D., Fredriksson, R., Kapa, I., Schioth, H.B. and Klovins, J. (2007) Formation of new genes explains lower intron density in mammalian Rhodopsin G protein-coupled receptors. *Mol Phylogenet Evol* 43, 864-80.
- [22] Bryson-Richardson, R.J., Logan, D.W., Currie, P.D. and Jackson, I.J. (2004) Large-scale analysis of gene structure in rhodopsin-like GPCRs: evidence for widespread loss of an ancient intron. *Gene* 338, 15-23.
- [23] Cardoso, J.C., Pinto, V.C., Vieira, F.A., Clark, M.S. and Power, D.M. (2006) Evolution of secretin family GPCR members in the metazoa. *BMC Evol Biol* 6, 108.
- [24] O'Hara, P.J., Sheppard, P.O., Thogersen, H., Venezia, D., Haldeman, B.A., McGrane, V., Houamed, K.M., Thomsen, C., Gilbert, T.L. and Mulvihill, E.R. (1993) The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* 11, 41-52.
- [25] Kuang, D., Yao, Y., Maclean, D., Wang, M., Hampson, D.R. and Chang, B.S. (2006) Ancestral reconstruction of the ligand-binding pocket of Family C G protein-coupled receptors. *Proc Natl Acad Sci U S A* 103, 14050-5.
- [26] Niswender, C.M. and Conn, P.J. Metabotropic glutamate receptors: physiology, pharmacology, and disease. *Annu Rev Pharmacol Toxicol* 50, 295-322.
- [27] Lin, H.H., Chang, G.W., Davies, J.Q., Stacey, M., Harris, J. and Gordon, S. (2004) Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *J Biol Chem* 279, 31823-32.
- [28] King, N., Hittinger, C.T. and Carroll, S.B. (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301, 361-3.
- [29] King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Sequencing, J.G., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V. and Rokhsar, D. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783-8.
- [30] Kulkarni, R.D., Thon, M.R., Pan, H. and Dean, R.A. (2005) Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biol* 6, R24.
- [31] Gookin, T.E., Kim, J. and Assmann, S.M. (2008) Whole proteome identification of plant candidate G-protein coupled receptors in *Arabidopsis*, rice, and poplar: computational prediction and in-vivo protein coupling. *Genome Biol* 9, R120.
- [32] Lelianova, V.G., Davletov, B.A., Sterling, A., Rahman, M.A., Grishin, E.V., Totty, N.F. and Ushkaryov, Y.A. (1997) Alpha-latrotoxin receptor, latrophilin, is a novel member of the secretin family of G protein-coupled receptors. *J Biol Chem* 272, 21504-8.
- [33] Little, K.D., Hemler, M.E. and Stipp, C.S. (2004) Dynamic regulation of a GPCR-tetraspanin-G protein complex on intact cells: central role of CD81 in facilitating GPR56-Galpha q/11 association. *Mol Biol Cell* 15, 2375-87.
- [34] Bjarnadottir, T.K., Fredriksson, R., Hoglund, P.J., Gloriam, D.E., Lagerstrom, M.C. and Schioth, H.B. (2004) The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics* 84, 23-33.

- [35] Stacey, M., Lin, H.H., Gordon, S. and McKnight, A.J. (2000) LNB-TM7, a group of seven-transmembrane proteins related to family-B G-protein-coupled receptors. *Trends Biochem Sci* 25, 284-9.
- [36] Kwakkenbos, M.J., Kop, E.N., Stacey, M., Matmati, M., Gordon, S., Lin, H.H. and Hamann, J. (2004) The EGF-TM7 family: a postgenomic view. *Immunogenetics* 55, 655-66.
- [37] Fredriksson, R., Gloriam, D.E., Hoglund, P.J., Lagerstrom, M.C. and Schioth, H.B. (2003) There exist at least 30 human G-protein-coupled receptors with long Ser/Thrich N-termini. *Biochem Biophys Res Commun* 301, 725-34.
- [38] Hadjantonakis, A.K., Formstone, C.J. and Little, P.F. (1998) mCelsr1 is an evolutionarily conserved seven-pass transmembrane receptor and is expressed during mouse embryonic development. *Mech Dev* 78, 91-5.
- [39] Formstone, C.J. and Little, P.F. (2001) The flamingo-related mouse Celsr family (Celsr1-3) genes exhibit distinct patterns of expression during embryonic development. *Mech Dev* 109, 91-4.
- [40] Usui, T., Shima, Y., Shimada, Y., Hirano, S., Burgess, R.W., Schwarz, T.L., Takeichi, M. and Uemura, T. (1999) Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Frizzled. *Cell* 98, 585-95.
- [41] Lawrence, P.A., Struhl, G. and Casal, J. (2007) Planar cell polarity: one or two pathways? *Nat Rev Genet* 8, 555-63.
- [42] Strutt, H. and Strutt, D. (2008) Differential stability of flamingo protein complexes underlies the establishment of planar polarity. *Curr Biol* 18, 1555-64.
- [43] Sudhof, T.C. (2001) alpha-Latrotoxin and its receptors: neurexins and CIRL/latrophilins. *Annu Rev Neurosci* 24, 933-62.
- [44] Langenhan, T., Promel, S., Mestek, L., Esmaeili, B., Waller-Evans, H., Hennig, C., Kohara, Y., Avery, L., Vakonakis, I., Schnabel, R. and Russ, A.P. (2009) Latrophilin signaling links anterior-posterior tissue polarity and oriented cell divisions in the *C. elegans* embryo. *Dev Cell* 17, 494-504.
- [45] Rovati, G.E., Capra, V. and Neubig, R.R. (2007) The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state. *Mol Pharmacol* 71, 959-64.
- [46] Markovic, D. and Grammatopoulos, D.K. (2009) Focus on the splicing of secretin GPCRs transmembrane-domain 7. *Trends Biochem Sci* 34, 443-52.
- [47] Parthier, C., Kleinschmidt, M., Neumann, P., Rudolph, R., Manhart, S., Schlenzig, D., Fanghanel, J., Rahfeld, J.U., Demuth, H.U. and Stubbs, M.T. (2007) Crystal structure of the incretin-bound extracellular domain of a G protein-coupled receptor. *Proc Natl Acad Sci U S A* 104, 13942-7.
- [48] Harmor, A.J. (2001) Family-B G-protein-coupled receptors. *Genome Biol* 2, REVIEWS3013.
- [49] Adell, T., Nefkens, I. and Muller, W.E. (2003) Polarity factor 'Frizzled' in the demosponge *Suberites domuncula*: identification, expression and localization of the receptor in the epithelium/pinacoderm(1). *FEBS Lett* 554, 363-8.
- [50] Adell, T., Thakur, A.N. and Muller, W.E. (2007) Isolation and characterization of Wnt pathway-related genes from Porifera. *Cell Biol Int* 31, 939-49.
- [51] Philipp, I., Aufschnaiter, R., Ozbek, S., Pontasch, S., Jenewein, M., Watanabe, H., Rentzsch, F., Holstein, T.W. and Hobmayer, B. (2009) Wnt/beta-catenin and noncanonical Wnt signaling interact in tissue evagination in the simple eumetazoan *Hydra*. *Proc Natl Acad Sci U S A* 106, 4290-5.
- [52] Momose, T. and Houlston, E. (2007) Two oppositely localised frizzled RNAs as axis determinants in a cnidarian embryo. *PLoS Biol* 5, e70.
- [53] Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev,

- I.V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q. and Rokhsar, D.S. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317, 86-94.
- [54] Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I.V., Buss, L.W., Schierwater, B., Dellaporta, S.L. and Rokhsar, D.S. (2008) The Trichoplax genome and the nature of placozoans. *Nature* 454, 955-60.
- [55] Prabhu, Y. and Eichinger, L. (2006) The Dictyostelium repertoire of seven transmembrane domain receptors. *Eur J Cell Biol* 85, 937-46.
- [56] Schoneberg, T., Schulz, A. and Gudermann, T. (2002) The structural basis of G-protein-coupled receptor function and dysfunction in human diseases. *Rev Physiol Biochem Pharmacol* 144, 143-227.
- [57] Nothacker, H.P. and Grimmelikhuijzen, C.J. (1993) Molecular cloning of a novel, putative G protein-coupled receptor from sea anemones structurally related to members of the FSH, TSH, LH/CG receptor family from mammals. *Biochem Biophys Res Commun* 197, 1062-9.
- [58] Kudo, M., Chen, T., Nakabayashi, K., Hsu, S.Y. and Hsueh, A.J. (2000) The nematode leucine-rich repeat-containing, G protein-coupled receptor (LGR) protein homologous to vertebrate gonadotropin and thyrotropin receptors is constitutively active in mammalian cells. *Mol Endocrinol* 14, 272-84.
- [59] Tensen, C.P., Van Kesteren, E.R., Planta, R.J., Cox, K.J., Burke, J.F., van Heerikhuizen, H. and Vreugdenhil, E. (1994) A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role for lipoproteins in G-linked signal transduction. *Proc Natl Acad Sci U S A* 91, 4816-20.
- [60] Kajava, A.V. (1998) Structural diversity of leucine-rich repeat proteins. *J Mol Biol* 277, 519-27.
- [61] Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M. and Zwiebel, L.J. (2002) G protein-coupled receptors in *Anopheles gambiae*. *Science* 298, 176-8.
- [62] Dehal, P. and Boore, J.L. (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3, e314.
- [63] Kuraku, S., Meyer, A. and Kuratani, S. (2009) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol Biol Evol* 26, 47-59.
- [64] Woods, I.G., Wilson, C., Friedlander, B., Chang, P., Reyes, D.K., Nix, R., Kelly, P.D., Chu, F., Postlethwait, J.H. and Talbot, W.S. (2005) The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res* 15, 1307-14.
- [65] Noonan, J.P., Grimwood, J., Danke, J., Schmutz, J., Dickson, M., Amemiya, C.T. and Myers, R.M. (2004) Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. *Genome Res* 14, 2397-405.
- [66] Semyonov, J., Park, J.I., Chang, C.L. and Hsu, S.Y. (2008) GPCR genes are preferentially retained after whole genome duplication. *PLoS One* 3, e1903.
- [67] Perez, D.M. (2003) The evolutionarily triumphant G-protein-coupled receptor. *Mol Pharmacol* 63, 1202-5.
- [68] Krautwurst, D. (2008) Human olfactory receptor families and their odorants. *Chem Biodivers* 5, 842-52.
- [69] Rouquier, S. and Giorgi, D. (2007) Olfactory receptor gene repertoires in mammals. *Mutat Res* 616, 95-102.
- [70] Kratz, E., Dugas, J.C. and Ngai, J. (2002) Odorant receptor gene regulation: implications from genomic organization. *Trends Genet* 18, 29-34.

- [71] Zlotnik, A., Yoshie, O. and Nomiya, H. (2006) The chemokine and chemokine receptor superfamilies and their molecular evolution. *Genome Biol* 7, 243.
- [72] Le Crom, S., Kapsimali, M., Barome, P.O. and Vernier, P. (2003) Dopamine receptors for every species: gene duplications and functional diversification in Craniates. *J Struct Funct Genomics* 3, 161-76.
- [73] Gloriam, D.E., Bjarnadottir, T.K., Yan, Y.L., Postlethwait, J.H., Schioth, H.B. and Fredriksson, R. (2005) The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish. *Mol Phylogenet Evol* 35, 470-82.
- [74] Hashiguchi, Y. and Nishida, M. (2007) Evolution of trace amine associated receptor (TAAR) gene family in vertebrates: lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium. *Mol Biol Evol* 24, 2099-107.
- [75] Grus, W.E., Shi, P. and Zhang, J. (2007) Largest vertebrate vomeronasal type 1 receptor gene repertoire in the semiaquatic platypus. *Mol Biol Evol* 24, 2153-7.
- [76] Schoneberg, T., Hermsdorf, T., Engemaier, E., Engel, K., Liebscher, I., Thor, D., Zierau, K., Rompler, H. and Schulz, A. (2007) Structural and functional evolution of the P2Y(12)-like receptor group. *Purinergic Signal* 3, 255-68.
- [77] Romero, D. and Palacios, R. (1997) Gene amplification and genomic plasticity in prokaryotes. *Annu Rev Genet* 31, 91-111.
- [78] Cohen, S., Agmon, N., Yacobi, K., Mislovati, M. and Segal, D. (2005) Evidence for rolling circle replication of tandem genes in *Drosophila*. *Nucleic Acids Res* 33, 4519-26.
- [79] Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 18, 486.
- [80] Staubert, C., Boselt, I., Bohnkamp, J., Rompler, H., Enard, W. and Schoneberg, T. Structural and functional evolution of the trace amine-associated receptors TAAR3, TAAR4 and TAAR5 in primates. *PLoS One* 5, e11133.
- [81] Rompler, H., Schulz, A., Pitra, C., Coop, G., Przeworski, M., Paabo, S. and Schoneberg, T. (2005) The rise and fall of the chemoattractant receptor GPR33. *J Biol Chem* 280, 31068-75.
- [82] Gaillard, I., Rouquier, S., Chavanieu, A., Mollard, P. and Giorgi, D. (2004) Amino-acid changes acquired during evolution by olfactory receptor 912-93 modify the specificity of odorant recognition. *Hum Mol Genet* 13, 771-80.
- [83] Rompler, H., Yu, H.T., Arnold, A., Orth, A. and Schoneberg, T. (2006) Functional consequences of naturally occurring DRY motif variants in the mammalian chemoattractant receptor GPR33. *Genomics* 87, 724-32.
- [84] Mombaerts, P. (2001) The human repertoire of odorant receptor genes and pseudogenes. *Annu Rev Genomics Hum Genet* 2, 493-510.
- [85] Young, J.M. and Trask, B.J. (2002) The sense of smell: genomics of vertebrate odorant receptors. *Hum Mol Genet* 11, 1153-60.
- [86] Go, Y., Satta, Y., Takenaka, O. and Takahata, N. (2005) Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics* 170, 313-26.
- [87] Niimura, Y. and Nei, M. (2006) Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J Hum Genet* 51, 505-17.
- [88] Liman, E.R. and Innan, H. (2003) Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc Natl Acad Sci U S A* 100, 3328-32.
- [89] Wang, X., Thomas, S.D. and Zhang, J. (2004) Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum Mol Genet* 13, 2671-8.



- [90] Zhang, J. and Webb, D.M. (2003) Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc Natl Acad Sci U S A* 100, 8337-41.
- [91] Schoneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Rompler, H. and Sangkuhl, K. (2004) Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol Ther* 104, 173-206.
- [92] Barbaric, I., Miller, G. and Dear, T.N. (2007) Appearances can be deceiving: phenotypes of knockout mice. *Brief Funct Genomic Proteomic* 6, 91-103.
- [93] Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., J, J.S., Adams, M.D. and Cargill, M. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* 3, e170.
- [94] Schulz, A. and Schoneberg, T. (2003) The structural evolution of a P2Y-like G-protein-coupled receptor. *J Biol Chem* 278, 35531-41.
- [95] Rouquier, S., Blancher, A. and Giorgi, D. (2000) The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. *Proc Natl Acad Sci U S A* 97, 2870-4.
- [96] Gilad, Y., Przeworski, M. and Lancet, D. (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2, E5.
- [97] Kishida, T., Kubota, S., Shirayama, Y. and Fukami, H. (2007) The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biol Lett* 3, 428-30.
- [98] Bamshad, M. and Wooding, S.P. (2003) Signatures of natural selection in the human genome. *Nat Rev Genet* 4, 99-111.
- [99] Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y. et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851-61.
- [100] Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Matukumalli, L.K., McEwan, J.C., Nazareth, L.V., Schnabel, R.D., Weinstock, G.M., Wheeler, D.A., Ajmone-Marsan, P., Boettcher, P.J., Caetano, A.R., Garcia, J.F., Hanotte, O., Mariani, P., Skow, L.C., Sonstegard, T.S., Williams, J.L., Diallo, B., Hailemariam, L., Martinez, M.L., Morris, C.A., Silva, L.O., Spelman, R.J., Mulatu, W., Zhao, K., Abbey, C.A., Agaba, M., Araujo, F.R., Bunch, R.J., Burton, J., Gorni, C., Olivier, H., Harrison, B.E., Luff, B., Machado, M.A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Thomas, M.B., Valentini, A., Williams, P., Womack, J., Woolliams, J.A., Liu, Y., Qin, X., Worley, K.C., Gao, C., Jiang, H., Moore, S.S., Ren, Y., Song, X.Z., Bustamante, C.D., Hernandez, R.D., Muzny, D.M., Patil, S., San Lucas, A., Fu, Q., Kent, M.P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J.J., Murdoch, B.,

- Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M.E., Hayes, B.J., Bradley, D.G., Barbosa da Silva, M., Lau, L.P., Liu, G.E., Lynn, D.J., Panzitta, F. and Dodds, K.G. (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528-32.
- [101] Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. and Bustamante, C. (2005) Genomic scans for selective sweeps using SNP data. *Genome Res* 15, 1566-75.
- [102] Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D. and Lander, E.S. (2006) Positive natural selection in the human lineage. *Science* 312, 1614-20.
- [103] Lin, Y.J., Seroude, L. and Benzer, S. (1998) Extended life-span and stress resistance in the *Drosophila* mutant methuselah. *Science* 282, 943-6.
- [104] Schmidt, P.S., Duvernell, D.D. and Eanes, W.F. (2000) Adaptive evolution of a candidate gene for aging in *Drosophila*. *Proc Natl Acad Sci U S A* 97, 10861-5.
- [105] Manolio, T.A. and Collins, F.S. (2009) The HapMap and genome-wide association studies in diagnosis and therapy. *Annu Rev Med* 60, 443-56.
- [106] Hunter, K.W. and Crawford, N.P. (2008) The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. *Annu Rev Genet* 42, 131-41.
- [107] Liu, X.G., Tan, L.J., Lei, S.F., Liu, Y.J., Shen, H., Wang, L., Yan, H., Guo, Y.F., Xiong, D.H., Chen, X.D., Pan, F., Yang, T.L., Zhang, Y.P., Guo, Y., Tang, N.L., Zhu, X.Z., Deng, H.Y., Levy, S., Recker, R.R., Papasian, C.J. and Deng, H.W. (2009) Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. *Am J Hum Genet* 84, 418-23.
- [108] Tonjes, A., Koriath, M., Schleinitz, D., Dietrich, K., Bottcher, Y., Rayner, N.W., Almgren, P., Enigk, B., Richter, O., Rohm, S., Fischer-Rosinsky, A., Pfeiffer, A., Hoffmann, K., Krohn, K., Aust, G., Spranger, J., Groop, L., Bluher, M., Kovacs, P. and Stumvoll, M. (2009) Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs. *Hum Mol Genet* 18, 4662-8.
- [109] Moreno-Estrada, A., Casals, F., Ramirez-Soriano, A., Oliva, B., Calafell, F., Bertranpetit, J. and Bosch, E. (2008) Signatures of selection in the human olfactory receptor OR5I1 gene. *Mol Biol Evol* 25, 144-54.
- [110] Shi, P., Zhang, J., Yang, H. and Zhang, Y.P. (2003) Adaptive diversification of bitter taste receptor genes in Mammalian evolution. *Mol Biol Evol* 20, 805-14.
- [111] Soranzo, N., Bufe, B., Sabeti, P.C., Wilson, J.F., Weale, M.E., Marguerie, R., Meyerhof, W. and Goldstein, D.B. (2005) Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. *Curr Biol* 15, 1257-65.
- [112] Hofreiter, M. and Schoneberg, T. The genetic and evolutionary basis of colour variation in vertebrates. *Cell Mol Life Sci* 67, 2591-603.
- [113] Harding, R.M., Healy, E., Ray, A.J., Ellis, N.S., Flanagan, N., Todd, C., Dixon, C., Sajantila, A., Jackson, I.J., Birch-Machin, M.A. and Rees, J.L. (2000) Evidence for variable selective pressures at MC1R. *Am J Hum Genet* 66, 1351-61.
- [114] Rees, J.L. (2000) The melanocortin 1 receptor (MC1R): more than just red hair. *Pigment Cell Res* 13, 135-40.
- [115] John, P.R., Makova, K., Li, W.H., Jenkins, T. and Ramsay, M. (2003) DNA polymorphism and selection at the melanocortin-1 receptor gene in normally pigmented southern African individuals. *Ann N Y Acad Sci* 994, 299-306.
- [116] Lalueza-Fox, C., Rompler, H., Caramelli, D., Staubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schoneberg, T., Bertranpetit, J. and Hofreiter, M. (2007) A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453-5.

- [117] Rompler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schoneberg, T. and Hofreiter, M. (2006) Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313, 62.
- [118] Nachman, M.W., Hoekstra, H.E. and D'Agostino, S.L. (2003) The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci U S A* 100, 5268-73.
- [119] Boselt, I., Rompler, H., Hermsdorf, T., Thor, D., Busch, W., Schulz, A. and Schoneberg, T. (2009) Involvement of the V2 vasopressin receptor in adaptation to limited water supply. *PLoS One* 4, e5573.
- [120] Hoekstra, H.E., Hirschmann, R.J., Bunday, R.A., Insel, P.A. and Crossland, J.P. (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101-4.
- [121] Steiner, C.C., Rompler, H., Boettger, L.M., Schoneberg, T. and Hoekstra, H.E. (2009) The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Mol Biol Evol* 26, 35-45.
- [122] Rosenblum, E.B. (2006) Convergent evolution and divergent selection: lizards at the White Sands ecotone. *Am Nat* 167, 1-15.
- [123] Rosenblum, E.B., Hoekstra, H.E. and Nachman, M.W. (2004) Adaptive reptile color variation and the evolution of the *Mc1r* gene. *Evolution* 58, 1794-808.
- [124] Hamblin, M.T. and Di Rienzo, A. (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* 66, 1669-79.
- [125] Hadley, T.J. and Peiper, S.C. (1997) From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. *Blood* 89, 3077-91.
- [126] Liberles, S.D. and Buck, L.B. (2006) A second class of chemosensory receptors in the olfactory epithelium. *Nature* 442, 645-50.
- [127] Voight, B.F., Kudravalli, S., Wen, X. and Pritchard, J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4, e72.
- [128] Arbiza, L., Dopazo, J. and Dopazo, H. (2006) Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLoS Comput Biol* 2, e38.
- [129] Tang, K., Thornton, K.R. and Stoneking, M. (2007) A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biol* 5, e171.
- [130] Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Wayne, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L. et al. (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913-8.
- [131] Williamson, S.H., Hubisz, M.J., Clark, A.G., Payseur, B.A., Bustamante, C.D. and Nielsen, R. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet* 3, e90.

- [132] Barreiro, L.B., Laval, G., Quach, H., Patin, E. and Quintana-Murci, L. (2008) Natural selection has driven population differentiation in modern humans. *Nat Genet* 40, 340-5.

## Tables

**Table 1: Potentially selected GPCR genes in the evolution of modern human populations identified in genome-wide studies.**

The following methods, mainly testing for positive selection, were applied:  $K_a/K_s$ : rate of non-synonymous to the rate of synonymous substitutions; CLR: composite likelihood ratio; EHH: extended haplotype homozygosity; iHS: integrated haplotype score; Rsb: relative integrated EHH of a site between populations; XP-EHH: cross population EHH; LRH: long-range haplotype;  $F_{ST}$ : fixation index; \*: selection candidates as determined by two or more different detection methods;  $\Psi$ : pseudogenes

GPCR genes	Data source	Detection Method	Reference
LGR8, CYSLTR2, OR2B2, OR511, OR2W1	Gene coding regions of human and chimpanzee	$K_a/K_s$	[93]
CXCR4*, GPR73	SNP data from HapMap1 (analysis restricted to chromosome 2)	CLR	[101]
Africans: GPR34, GPR82, OR4C2P $\Psi$ , OR4C3, OR4C4P $\Psi$ , OR4C5 $\Psi$ , OR5T2, OR8J2 $\Psi$ , OR8K5, OR8V1P $\Psi$ , OR10V2P $\Psi$ , OR10Y1P $\Psi$ Europeans: OR2AK2, OR2L1P $\Psi$ , OR2L2, OR2L5 $\Psi$ , OR2L9P $\Psi$ , OR2L13 Asians: GRM3*, GRM8*	SNP data from HapMap1	iHS (EHH based)	[127]
GRM6, TAS1R1, GPR111*, PTGER4, OR2A14, OR5D18, OR51D1, OR52W1	Gene coding regions of human and chimpanzee	$K_a/K_s$	[128]
Europeans: CXCR4*, CXCR7, NPY1R, GPR65, GPR151, DRD5 Asians: BAI3*, GRM3*, GPR65	SNP data from HapMap2 and Perlegen Sciences project	Rsb (EHH based)	[129]
Europeans: GPR64* Asians: CELSR1*	SNP data from HapMap2	XP-EHH	[130]
Africans: BAI3*, OR52A1, OR52A4, OR52A5, OR4K1, OR4K5 Europeans: GPR64*, GPR177, TACR3, OR4K1, OR4K5 Asians: GPR111*, GPR115, OR4A47*	SNP data from HapMap2	LRH and/or iHS (EHH based)	[99]
Africans: OR4P4 Asians: OR9K2, MC2R	SNP data from Perlegen Sciences project	CLR	[131]
PTGER3, GRM7, GRM8*, LGR7, BAI3*, P2RY2, ADORA2B, CELSR1*, OR52K2	SNP data from HapMap2	$F_{ST}$	[132]