The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity

Masatoshi Nei, Yoshihito Niimura‡ and Masafumi Nozawa**

Abstract | Chemosensory receptors are essential for the survival of organisms that range from bacteria to mammals. Recent studies have shown that the numbers of functional chemosensory receptor genes and pseudogenes vary enormously among the genomes of different animal species. Although much of the variation can be explained by the adaptation of organisms to different environments, it has become clear that a substantial portion is generated by genomic drift, a random process of gene duplication and deletion. Genomic drift also generates a substantial amount of copy-number variation in chemosensory receptor genes within species. It seems that mutation by gene duplication and inactivation has important roles in both the adaptive and non-adaptive evolution of chemosensation.

Chemosensation
The sense of smell and taste.

Multigene family
A group of genes that have descended from a common ancestor, and therefore have similar functions and similar DNA sequences.

**Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, 328 Mueller Laboratory, University Park, Pennsylvania 16802, USA.*
‡Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan.
Correspondence to M.N.
e-mail: nxm2@psu.edu
doi:10.1038/nrg2480
Published online
11 November 2008

In animals, chemosensory receptors (CRs) are used to find food, detect mates and offspring, recognize territories and avoid danger. Animal genomes contain a large number of CR genes that allow these species to distinguish between a myriad of odour and taste chemicals. Recent publications of the whole genome sequences of various organisms have made it possible to identify all of the CR genes[1–3] and study their evolution. These studies have revealed that the numbers of these genes are very large and vary enormously among different species and that each genome contains a surprisingly large number of CR pseudogenes. There are also significant numbers of copy-number variations in the CR genes within species. To explain these new findings, various evolutionary theories have been proposed[4–8]. Some interspecific variations in copy number can readily be explained by the adaptation of organisms to different environments, but adaptation theory does not necessarily account for other aspects of changes, such as dramatic increases in the number of pseudogenes, or extensive gains and losses of genes during evolutionary processes. It is a challenge to develop a comprehensive theory that explains both interspecific and intraspecific variation in gene copy numbers in various groups of animals, particularly because only a limited number of species have been studied and our understanding of the biochemical basis of chemosensation is incomplete. Here, we provide an overview of the general features of the evolution of CR genes that have emerged from studies in recent years.

Receptor gene families and their functions

Vertebrate chemosensory receptors. The vertebrate CRs are encoded by six different multigene families: olfactory receptor (OR)[9], trace amine-associated receptor (TAAR)[10], vomeronasal receptor type 1 and 2 (V1R[11] and V2R[12,13]), and taste receptor type 1 and 2 (T1R[14] and T2R[15,16]) genes. The OR, TAAR, V1R and V2R genes encode olfactory or pheromone receptors, whereas the T1R and T2R genes encode taste receptors (TRs). All of the proteins encoded by the CR genes are G protein-coupled receptors (GPCRs) that have seven transmembrane α-helical regions (FIG. 1).

OR genes are predominantly expressed in sensory neurons of the main olfactory epithelium (MOE) in the nasal cavity. Mammals detect many types of chemicals in the air and some in the water as odorants, whereas fishes recognize water-soluble molecules, such as amino acids, bile acids, sex steroids and prostaglandins. Some mammalian OR genes are known to be expressed in other tissues, including the testis, tongue, brain and placenta[17]. However, the functional significance of such 'ectopic expression' of OR genes is not definitively known. TAARs are also expressed in the MOE. These receptors were first identified as brain receptors for the trace amines, a collection of amines that are present at low concentrations in the central nervous system[18]. TAARs were originally suspected to be involved in psychiatric disorders[19] but are now known to function as a second class of olfactory receptors[10]. Some mouse TAARs recognize volatile amines that are present in urine, and it seems that the

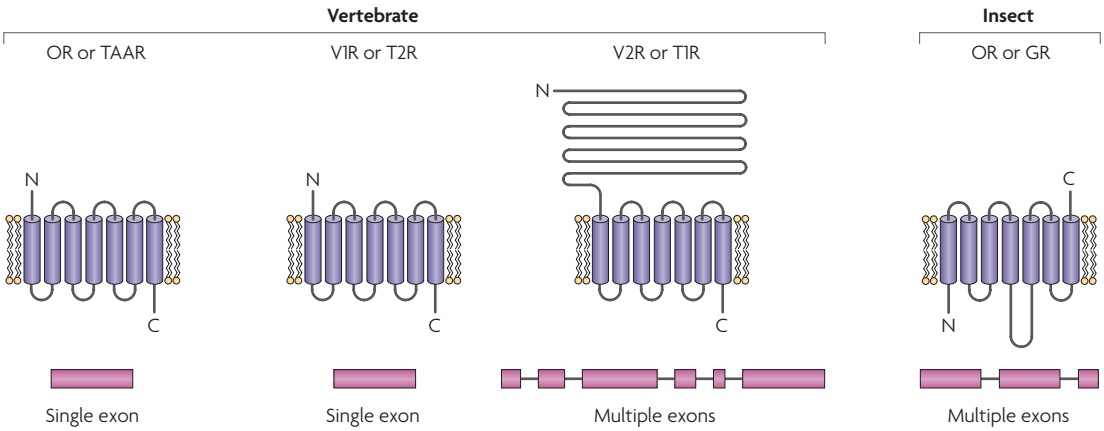


Figure 1 | **Chemosensory receptors and their genes.** Typical membrane topologies of chemosensory receptors, showing the orientation of the carboxyl terminus and the amino terminus, and the exon–intron structures of their genes. GR, gustatory receptor; OR, olfactory receptor; T1R, taste receptor type 1; T2R, taste receptor type 2; TAAR, trace amine-associated receptor; V1R, vomeronasal receptor type 1; V2R, vomeronasal receptor type 2.

TAARs function to detect ligands associated with social cues[10]. TAARs share sequence similarity with ORs, and both OR and TAAR genes lack introns in their coding regions (FIG. 1).

Most mammals possess an additional olfactory organ called the vomeronasal organ (VNO). The VNO is located in a region that is proximal to the vomer bone in the nasal cavity, and is distinct from the MOE, in which most of the OR genes are expressed. Mammalian vomeronasal receptor (VR) neurons express members of one of the two families of chemosensory receptors, V1Rs[11] and V2Rs[12,13], which are thought to be responsible for the detection of pheromones (see REFS 20,21 for reviews). The VNO was previously thought to be a specialized organ for pheromone detection, but it is now known that the VNO and MOE share some overlapping functions[22]. Like the OR genes, V1R genes have no introns, whereas the V2R genes are interrupted by introns and the V2Rs are characterized by a long extracellular amino-terminal tail (FIG. 1).

Two types of taste receptors, the T1Rs and T2Rs, are expressed in the taste buds of the tongue. There are five types of taste: sweet, sour, bitter, salt and umami (which means delicious in Japanese and corresponds to the taste of L-glutamate[23]). Of these five types of taste, salt and sour are detected by ion channels. By contrast, sweet and umami tastes are detected by T1Rs[14], and bitterness is recognized by T2Rs[15,16] (see REFS 24,25 for reviews). In general, mammals have only three T1R genes (*T1R1*, *T1R2* and *T1R3*). T1R2 and T1R3 form a heterodimer that functions as a sweet receptor that responds to a range of sweet substances, whereas the T1R1–T1R3 heterodimer acts as a umami receptor. T1Rs and T2Rs share significant sequence similarity with V2Rs and V1Rs, respectively (FIG. 1), whereas the ORs and TAARs, V1Rs and T2Rs, and V2Rs and T1Rs share almost no sequence similarity, despite their similar molecular structures.

Insect chemosensory receptors. Insects have only two different multigene families for CRs: ORs[26,27] and gustatory receptors (GRs)[28] (see REFS 29,30 for reviews). Insect OR and GR genes are distantly related and constitute a large superfamily of insect CR genes[31]. In *Drosophila* spp., the olfactory organs consist of the antenna and the maxillary palp on the head, whereas the gustatory organs are distributed on the entire body, including the proboscis, wings and legs[30]. ORs and GRs are primarily responsible for detecting odorants and tastants, respectively, but recent studies have suggested that some GRs recognize carbon dioxide[32] and pheromones[33]. There are common neuroanatomical features between the insect and vertebrate olfactory systems, but insect and vertebrate CR genes are strikingly different and share no sequence similarity[29]. Insect CRs have a seven-transmembrane region similar to vertebrate CRs, but their membrane topology is inverted compared with that of classic GPCRs, such that the N-terminal tail of the insect CR is found in the intracellular region[34] (FIG. 1). Furthermore, insect ORs function as heterodimers. The active receptor complex is formed by the heterodimerization of a member of the OR family with the ubiquitously expressed receptor Or83b[35].

Evolution of vertebrate receptors

OR genes. FIGURE 2a shows the numbers of functional genes and pseudogenes for CRs in various vertebrate species. All of the tetrapod animals that have been examined so far (including mammals, birds and amphibians) have 400–2,100 OR genes, but 20–50% of these are pseudogenes. These numbers are small compared with the number of potential odorants. However, ORs are known to detect odorants in a combinatorial manner[36]. Thus, a single OR may detect multiple odorants and a single odorant may be detected by multiple ORs. Therefore, millions of different odorants may be detected by a limited number of OR genes. There are a large number of OR subfamilies, which are loosely related to different types of odorants[37,38], but the detailed relationships between receptors and odorants remain unclear. In each species, the OR genes are distributed throughout the genome but usually exist as clusters of closely related genes[39] (FIG. 3).

Pheromone
A chemical substance that is released and detected by different individuals of the same species, and triggers physiological and behavioural responses.

Main olfactory epithelium
A specialized epithelial tissue in the nasal cavity in which olfactory receptor genes are expressed.

Ectopic expression
The expression of a gene in tissues other than the tissue in which it is normally expressed.

Vomeronasal organ
An auxiliary olfactory organ that is found in many mammals, reptiles and amphibians.

Proboscis
The tubular organ in insects that is used for feeding and sucking.

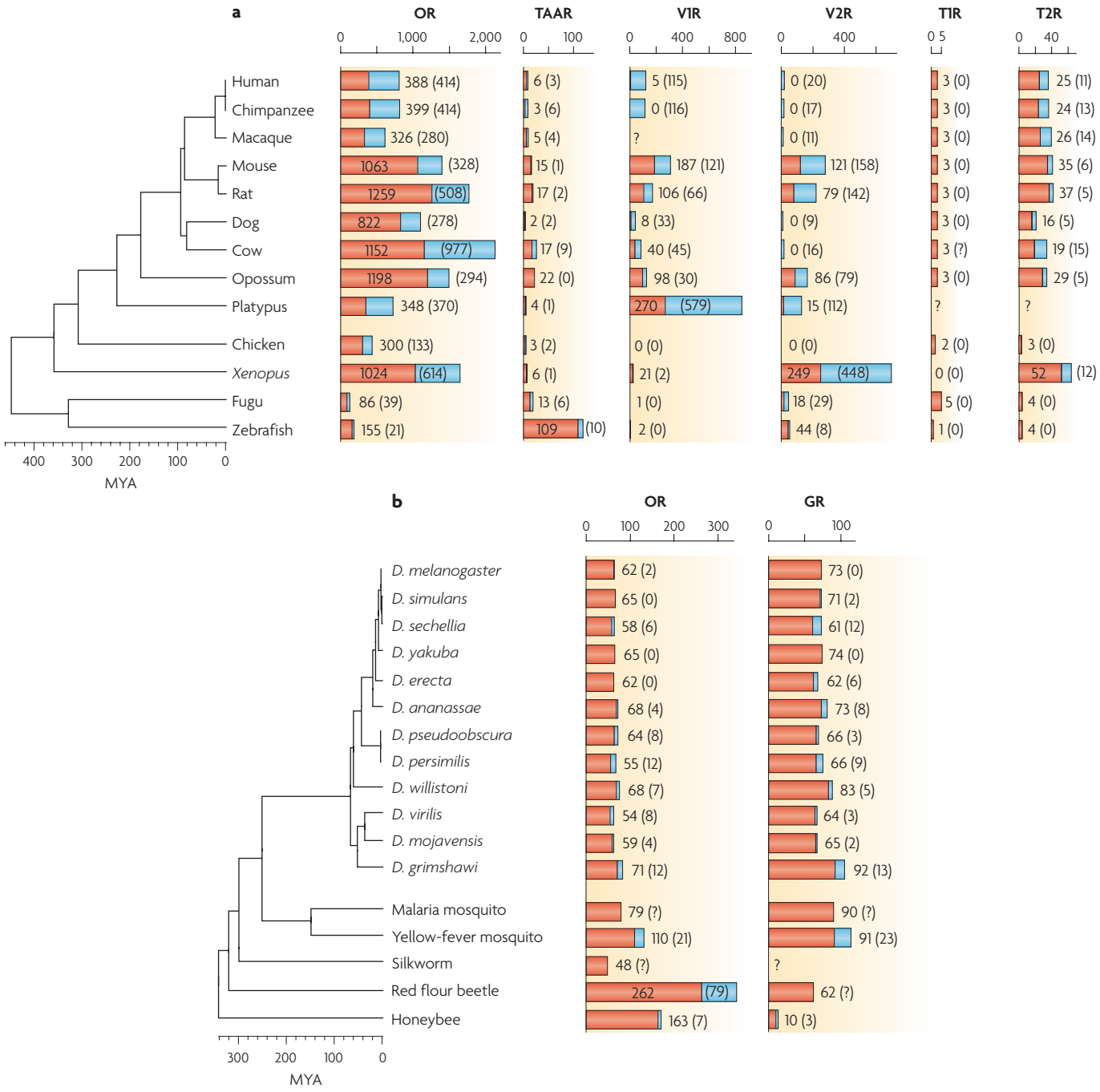
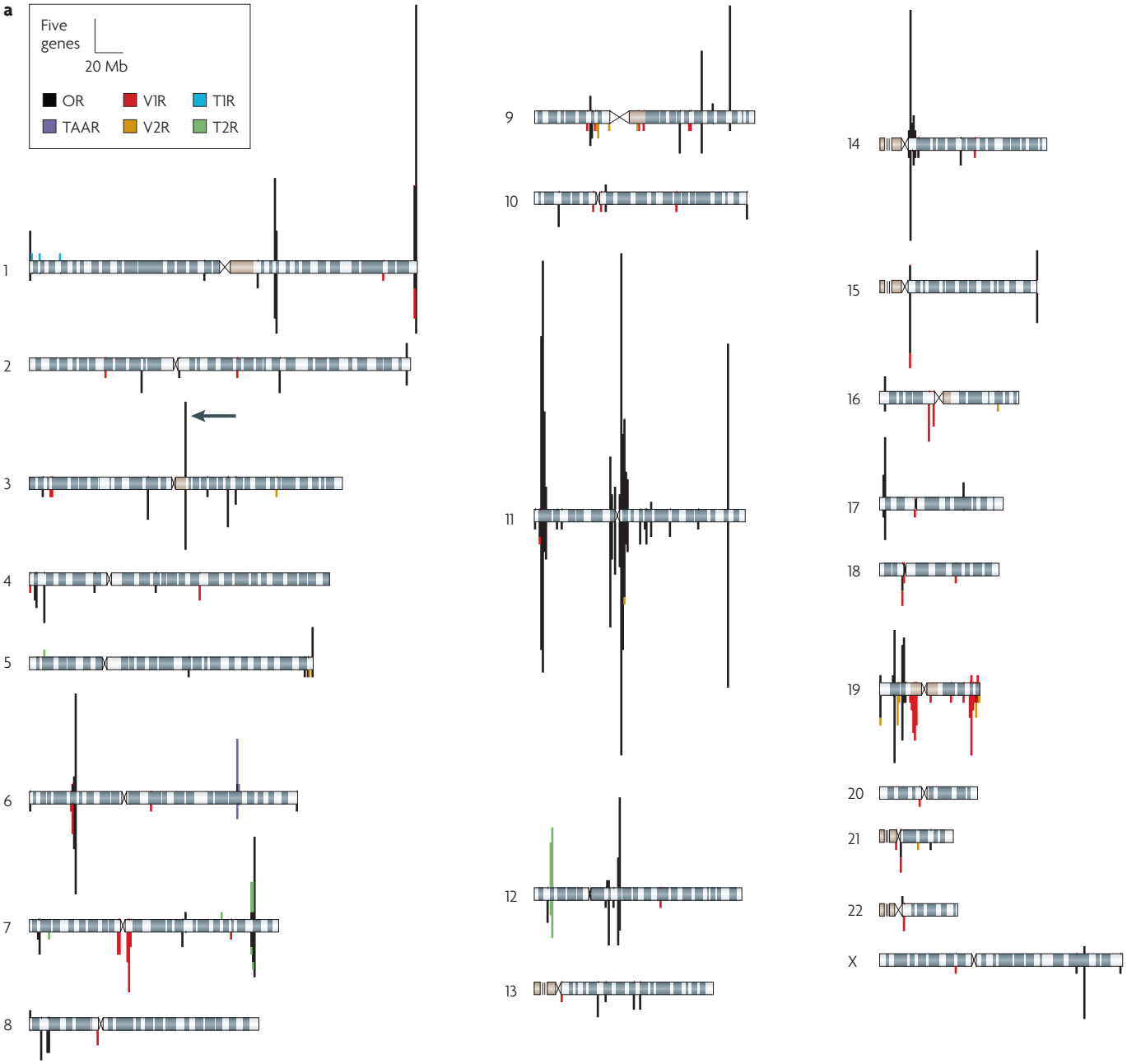


Figure 2 | **Numbers of chemosensory receptor genes in vertebrates and insects.** The red and blue bars represent the numbers of functional (intact) genes and pseudogenes (disrupted genes), respectively. Truncated genes, which are sequences that have a coding region without disruptive mutations but are apparently truncated owing to incomplete sequence assembly, are provisionally counted as functional genes. The numbers next to each bar represent the number of functional genes and the number of pseudogenes, which is shown in parentheses. A question mark indicates that data are unavailable. In part **a**, the OR data are from REFS 39,44,50 and Y.N. (unpublished observations), the TAAR data are from REFS 65,68,135 and M.N. (unpublished observations), the V1R data are from REFS 67,68,75, the V2R data are from REFS 68,69,75, the T1R data are from REF. 78 and Y. Go (personal communication), and the T2R data are from REF. 78 and Y. Go (personal communication). In part **b**, the *Drosophila* data are from REFS 81,85,136, the malaria mosquito data are from REFS 137,138, the yellow-fever mosquito data are from REFS 138,139, the silkworm data are from REF. 140, the red flour beetle data are from REFS 83,84 and the honeybee data are from REF. 82. The *Drosophila* spp. and mosquitoes belong to the order Diptera, and the remaining three species (silkworm, red flour beetle and honeybee) belong to different orders (Lepidoptera, Coleoptera and Hymenoptera, respectively). The *Xenopus* species referred to is the Western clawed frog, *Xenopus tropicalis*. GR, gustatory receptor; MYA, million years ago; OR, olfactory receptor; T1R, taste receptor type 1; T2R, taste receptor type 2; TAAR, trace amine-associated receptor; V1R, vomeronasal receptor type 1; V2R, vomeronasal receptor type 2.



◀ Figure 3 | **Distribution of chemosensory receptor genes in the human genome.** **a** | The vertical bars above and below the chromosomes represent the locations of functional genes and pseudogenes, respectively. The height of each bar indicates the number of genes present in a non-overlapping 1 Mb window. **b** | The gene content of the OR gene cluster indicated by the arrow in part **a**. The diagram (left) represents a 0.6 Mb region on chromosome 3. Ψ indicates a pseudogene. All of the OR genes are encoded on the same strand. The phylogenetic tree for the genes of this cluster (right) shows that the genes that are closely located to each other on a chromosome tend to be closely related evolutionarily. A human class I OR gene, *HsOR11.3.2*, was used as the out-group[39]. OR, olfactory receptor; T1R, taste receptor type 1; T2R, taste receptor type 2; TAAR, trace amine-associated receptor; V1R, vomeronasal receptor type 1; V2R, vomeronasal receptor type 2. Part **b** is modified, with permission, from REF. 39 © (2003) National Academy of Sciences.

OR genes are present in all vertebrate species. Several OR genes were identified in the lamprey, a primitive jawless vertebrate[40,41]. Fishes have approximately 100 OR genes, but this number is much smaller than the number of OR genes present in mammals (FIG. 2a). Phylogenetic analysis has shown that vertebrate OR genes can be classified into at least nine groups (α, β, γ, δ, ε, ζ, η, θ and κ), each of which originated from one or a few ancestral genes in the most recent common ancestor (MRCA) of fishes and tetrapods[6] (FIG. 4a). Interestingly, there was an enormous expansion in the number of genes in the α and γ groups in tetrapods, and these genes are often called the class I and II OR genes, respectively[41]. By contrast, the remaining groups of OR genes are present primarily in the fish and amphibian genomes. This observation suggests that the α and γ group genes function mostly to detect airborne odorants, and that the primary function of the remaining groups is to detect water-soluble odorants. It seems that the expansion of the group α and γ OR genes played important parts in the evolution of the teleost fishes into land vertebrates. Therefore, if we know the function of the groups of genes that expanded or contracted during the evolutionary process, it is easier to identify their adaptive significance.

However, it is not a simple task to uncover general principles of the evolution of OR genes that are applicable to a wide range of organisms. Primate species (human, chimpanzee and macaque) generally have a smaller number of OR genes than rodents, but the proportion of pseudogenes is higher in primates than in rodents (FIG. 2a). For example, mice have approximately 600 more genes than humans but have a much lower proportion of pseudogenes (approximately 24% compared with 52%)[4,42]. Why does this conspicuous difference in the proportion of pseudogenes in rodents and primates exist? A popular explanation is that hominoids and Old World monkeys are equipped with a complete trichromatic colour-vision system, and therefore the requirement for olfaction declined, which has resulted in a higher proportion of OR pseudogenes in hominoids and Old World monkeys (the vision-priority hypothesis)[5]. However, if we consider all of the placental mammals, this hypothesis does not hold, because the cow has dichromatic vision but still has a high proportion of pseudogenes (46%). Of course, one can argue that the proportion of pseudogenes is unimportant to test the vision-priority hypothesis, because the important issue is whether functional genes become pseudogenes with improved colour vision. If we accept this view and only consider the number of functional genes, the vision-priority hypothesis holds, because all three primate species have a substantially lower number of functional genes compared with other placental mammals.

The vision-priority hypothesis can be tested in another way by examining the extent of purifying selection in OR genes. Purifying selection is implied when the ratio (w) of the number of non-synonymous nucleotide substitutions per non-synonymous site (d_N) to the number of synonymous nucleotide substitutions per synonymous site (d_S) is smaller than 1 (REF. 43). When we compared 490 orthologous OR genes between mice and rats, the average value was 0.19 (M.N., unpublished observations), whereas the same ratio for 257 orthologous OR genes between humans and chimpanzees was 0.94 (REF. 44). Therefore, purifying selection has apparently been relaxed in humans and chimpanzees. This observation supports the vision-priority hypothesis. It is theoretically possible that the larger primate value of w is partly due to the difference in the effective population sizes of primates and rodents[45]. However, the difference between 0.19 and 0.94 ($P < 3 \times 10^{-6}$) seems too high to be due to this factor alone.

It is also interesting that both chickens and *Xenopus* spp. have excellent colour vision although the underlying genetic mechanism is different from that present in primates[46]. Although the quality of the genome sequence obtained from these species is currently low, the proportion of pseudogenes is also low (31% for chicken and 37% for *Xenopus* spp.; Y.N., unpublished observations). Both these observations and the presence of a well-developed olfactory system in birds[47] are not consistent with the vision-priority hypothesis. However, because there are major differences in the anatomy, physiology and lifestyle of chickens, *Xenopus* spp. and primates, it is unclear whether this type of study is meaningful. An alternative hypothesis called the brain-function hypothesis could explain the reduced number of OR genes in primates[48,49]. According to this hypothesis, higher brain function (such as good memory in primates) confers a greater olfactory ability than would be expected from the small number of OR genes. This hypothesis is attractive, but difficult to test.

The platypus also has a smaller number of functional OR genes and a larger fraction of pseudogenes (approximately 52%) than other mammals (FIG. 2a) but this observation can be explained in the following way[50]. Platypuses are semi-aquatic animals and have evolved from a land animal with OR genes that are primarily used to detect airborne odorants. They have a special sense in their bills, which combines electroreception and mechanoreception[51]. Platypuses can find prey with their eyes, ears and nostrils closed[51], and therefore many OR genes that were primarily used for airborne odorants may have become pseudogenes. The evolution of the platypus olfactory system is similar to that of toothed whales (including dolphins), in which the olfactory system deteriorated considerably when the echolocation

Non-synonymous nucleotide substitution
A nucleotide substitution that results in a change of an amino acid within the coding region of a gene.

Synonymous nucleotide substitution
A nucleotide substitution that does not change an amino acid in the coding region of a gene.

Electroreception
The ability of an animal to perceive electrical pulses.

Mechanoreception
The ability of an animal to detect certain kinds of stimuli, such as touch, sound and changes in atmospheric pressure or posture in its environment.

Toothed whales
A suborder of cetaceans that have teeth rather than baleens (plates of whalebone). Toothed whales include dolphins, sperm whales, beaked whales and killer whales.

Echolocation
A biological sonar mechanism used by several mammals, such as whales and bats. A high-pitched sound (usually clicks) is sent out by an animal, the sound bounces off an object and some of the sound returns to the animal. Whales perceive this returning echo to determine the shape, direction, distance and texture of the object.

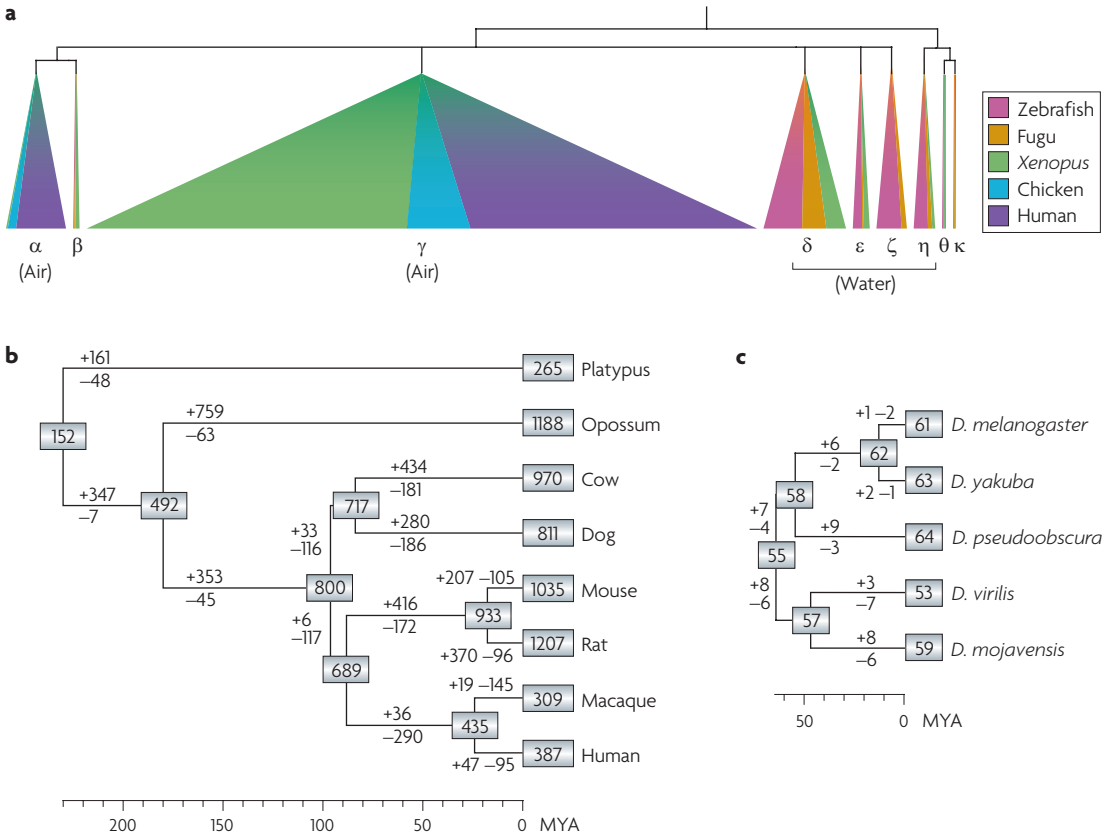


Figure 4 | **Evolutionary dynamics of olfactory receptor genes. a** | A phylogenetic tree of olfactory receptor (OR) genes from five vertebrate species. The genes that belong to different groups are represented by different coloured triangles. The size of each triangle is approximately proportional to the number of OR genes from each species. The colours that are obscured at the top of a triangle indicate that some gene duplications occurred before the divergence of the species. The root of the tree was determined using non-OR G protein-coupled receptor (GPCR) genes as the out-group. The α and γ group genes are proposed to primarily detect airborne odorants because they exist in tetrapods, whereas the δ, ε, ζ and η group genes that exist in fishes and *Xenopus* spp. appear to primarily detect water-soluble odorants. The functions of the group β, θ and κ genes are unclear. **b** | Evolutionary change of the number of OR genes in mammals. The numbers in rectangles represent the numbers of functional OR genes in the extant or ancestral species. The numbers with plus and minus signs indicate the numbers of gene gains and losses, respectively, for each branch. **c** | Evolutionary change of the number of OR genes in *Drosophila* spp. Truncated genes are excluded from the analysis in parts **b** and **c**. MYA, million years ago. Part **b** is modified from REF. 50. Part **c** is modified, with permission, from REF. 85 © (2007) National Academy of Sciences.

system evolved to allow adaptation to full aquatic life[52]. Recent studies have shown that the proportion of pseudogenes in toothed whales is high (more than 75%), although only a small number of genes have been examined[53–55].

Nevertheless, the proportion of pseudogenes is not always a good criterion for the study of the evolution of OR genes. Theoretically, the number of pseudogenes can rapidly increase through duplication because they are essentially neutral[56,57], but pseudogene copies should eventually become deleted or unidentifiable because of mutations. Furthermore, because functional genes are more important for adaptation, they should be retained in the genome longer than pseudogenes. However, the evolutionary changes that occur in functional OR genes or pseudogenes seem to be more complicated. Zhang *et al.*[58] recently reported that 67% of human OR pseudogenes and 80% of intact OR genes are expressed in the MOE. Some RNAs that are transcribed from pseudogenes have regulatory roles in gene expression[59–61]. In these cases, there is an unclear boundary between the definition of the intact genes and pseudogenes, and if transcripts from OR pseudogenes can also participate in regulation, we will need to study both intact OR genes and pseudogenes more carefully using gene-expression experiments. The human OR gene *OR1E3* (*OR17-210*), which was originally identified as a pseudogene owing to a two-nucleotide frameshift, was later found to be functional[62]. The presence of an initiation codon downstream of the frameshift mutation site allowed the *OR1E3* gene to encode a new protein sequence and acquire a function. The OR encoded by *OR1E3* lacks the first two transmembrane domains but contains a new domain at the carboxyl terminus. Because there are a large number of OR genes, there may be many other pseudogenes that have a function.

So far, we have only discussed the evolutionary change in gene copy number in the genome. In practice, it is important to identify the amino-acid changes that alter the function of OR genes, but this is a difficult problem to resolve because there are many gene copies and olfaction is determined in a combinatorial manner, as mentioned earlier. We discuss this complex problem for OR and other CR genes in a later section.

TAAR genes. The number of TAAR genes is generally smaller than the number of OR genes, except in zebrafish, and consequently the interspecific variation in TAAR gene copy number is also smaller than in OR genes (FIG. 2a). The reason for this low level of variation has not been well studied. However, gene families that have a low copy number throughout vertebrate species are expected to have some important biochemical function and to evolve slowly because of functional constraints. Indeed, in tetrapods, TAAR genes seem to be evolving at a slower rate in terms of amino-acid substitution compared with OR genes[63]. If TAAR genes are evolving more slowly, the degree of interspecific variation in these genes is expected to be small. However, the number of TAAR genes is unusually large in zebrafish[64,65]. A large number of TAAR genes have also been observed in several other teleost fishes, such as stickleback (64 genes) and medaka (32 genes)[65]. In these species, the copy number of a specific group of TAAR genes seems to have increased rapidly by tandem duplication[65]. However, the reason for the emergence of this group of genes is unclear.

VR genes. The numbers of V1R and V2R genes also vary extensively among mammalian orders[66–69] (FIG. 2a). Even if we consider only functional genes, the number of V1R genes varies from 0 (chimpanzee) to 270 (platypus), and the number of V2R genes varies from 0 (human, chimpanzee, macaque, dog and cow) to 121 (mouse). There is no clear relationship between the numbers of V1R and V2R genes. The numbers of V1R and V2R genes also do not seem to be correlated to the number of OR genes. The proportion of pseudogenes in the genome varies extensively, as for OR genes. In hominoids and Old World monkeys, the high proportion of V1R pseudogenes was initially thought to be related to the acquisition of trichromatic vision[70,71]. However, if we consider the variation in pseudogene number among all tetrapod species (FIG. 2a), it is difficult to relate variation to any specific environmental factor. There are six vertebrate species in which no functional V2R genes are observed, but all of these species, except chicken, have about a dozen or more pseudogenes. Theoretically, some pseudogenes may be transcribed, and the RNAs produced may have some gene regulatory functions, as mentioned earlier. However, we currently have no evidence to suggest that this theory is correct.

Humans have no functional V2R genes but have five intact V1R genes. It has been argued that although these five V1R genes have an open reading frame, they are not functional because a calcium channel gene (*TRPC2*) that is essential in the signal transduction pathway of the mouse VNO has become a pseudogene in the lineage that leads to hominoids and Old World monkeys[70,71]. However, at least one of the five V1R genes is expressed in the human olfactory mucosa[72]. Furthermore, a recent study suggests that these five genes can activate an OR-like signal transduction pathway in a heterologous expression system[73]. It is therefore possible that the products of these genes function as pheromone or olfactory receptors. Adult humans do not have a VNO but seem to be sensitive to pheromones[74]. Another interesting observation is that chickens have no functional or non-functional V1R and V2R genes or a VNO[75], although birds use pheromones for mate choice and other behaviours[76]. It is possible that some OR genes in the MOE are able to detect pheromones, as in humans[74,77].

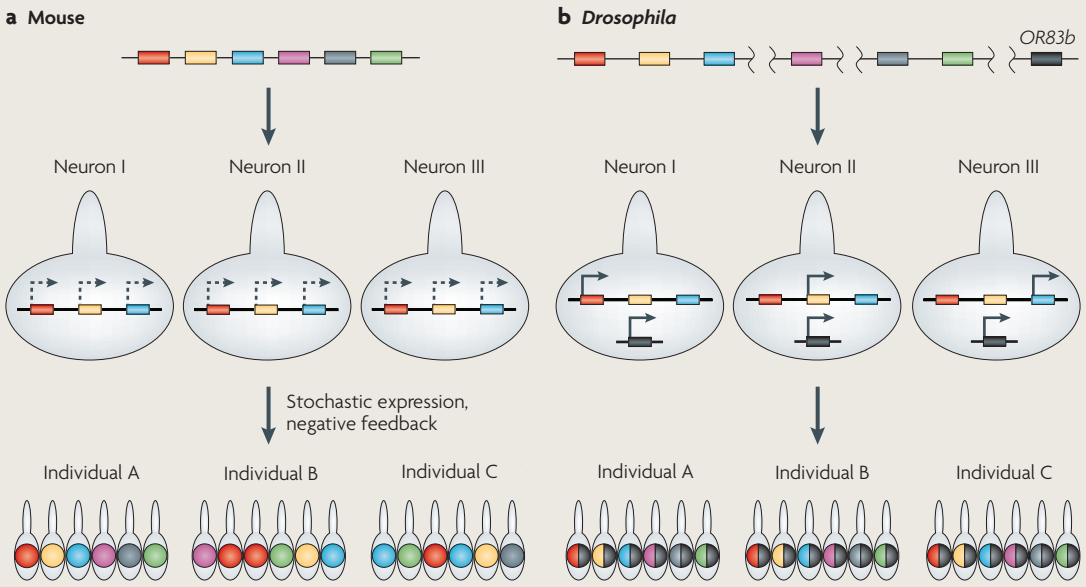
TR genes. The number of T2R (bitter taste receptor) genes is generally larger than the number of T1R (sweet and umami receptor) genes (FIG. 2a). This difference is probably due to the presence of various types of toxins in the environment and because a large number of T2R genes are required to protect animals from these toxins. The number of T2R genes also varies among different species, suggesting that different species encounter different toxins. The small number of T2R genes in dogs and cows could be explained by the fact that dogs are exclusively carnivorous and toxic substances are less common in animals than in plants. Dogs may therefore encounter toxic chemicals less frequently than omnivorous animals, and consequently may not need a large number of T2R genes[78]. Cows are ruminants and effectively carry out detoxification processes, and therefore the detection of dietary poisons might not be as important as in other mammals[78]. Compared with other organisms, *Xenopus* spp. have a large number of T2R genes.

By contrast, there are only three T1R genes (*T1R1, T1R2* and *T1R3*) in all of the mammalian species examined, except cats, in which *T1R2* is pseudogenized[79]. In other species, few or no T1R genes are present. The low number of T1R genes probably arises because each T1R is broadly tuned to detect various types of tastants. Phylogenetic analysis of T1R genes from all vertebrate species showed that the *T1R1, T1R2* and *T1R3* genes diverged before the separation of fishes and tetrapods[78,80]. During vertebrate evolution, gene duplication or deletion rarely occurred in the T1R gene family, except in a few species. Chickens and cats lost the *T1R2* gene, resulting in a loss of their ability to detect sweet tastes[78,79]. *Xenopus* spp. apparently lost all three T1R genes. By contrast, the number of T1R genes increased in some teleost fishes. Fugus have one *T1R1*, three *T1R2* and one *T1R3* gene, whereas sticklebacks have eight *T1R2* genes, in addition to one *T1R1* and one *T1R3* gene[80]. The evolutionary importance of the new duplicate *T1R2* genes in teleost fishes is currently unknown.

Evolution of insect chemosensory receptor genes

The number of CR genes in insects is much smaller than in vertebrates. *Drosophila melanogaster* has only 62 types of ORs that are encoded by 59 genes through alternative

Box 1 | **Models of olfactory receptor gene expression in mice and *Drosophila* spp.**



In mice (see figure, part **a**), olfactory receptor (OR) genes are present as genomic clusters that are located on different chromosomes. A single olfactory neuron expresses a single functional OR gene from a genomic region. It has been proposed that a functional OR gene is stochastically chosen to be expressed (broken arrows) in each olfactory neuron (only the first three OR genes are shown in each neuron), and the expression of that OR gene prevents the activation of other OR genes through negative feedback regulation[86,87]. Consequently, each individual has a different expression pattern of OR genes. According to this model, it is probable that the duplication or deletion of an OR gene does not affect the expression of any existing OR gene and therefore the number of OR genes could change easily during the evolutionary process. In *Drosophila* spp. (see figure, part **b**), OR genes are scattered throughout the genome and only some are clustered. It has been proposed that each olfactory neuron expresses one specific OR gene together with the ubiquitously expressed gene, *Or83b* (grey boxes). The specific OR gene tends to be expressed deterministically in a given olfactory neuron[29,88,89]. Therefore, the expression pattern of OR genes is essentially identical for different individuals. In this case, if an OR gene is duplicated or deleted, the gene expression pattern may be disturbed.

splicing, and 73 types of GRs that are encoded by a different set of 60 genes[31,81]. Interestingly, there are similar numbers of OR and GR genes in all of the insect species that have been studied so far, except for the red flour beetle and the honeybee (FIG. 2b). As for vertebrate OR genes, there is a considerable amount of interorder variation in the number of insect CR genes, and this variation seems to be partly generated by adaptation to different environmental conditions. For example, the honeybee has more than 160 intact OR genes but only 10 intact GR genes[82]. Most honeybee OR genes belong to a bee-specific subfamily. The occurrence of bee-specific gene expansion presumably facilitated the evolution of the remarkable olfactory abilities of the honeybee, including the recognition of diverse floral odours and complex pheromone blends, which allowed the coordination of caste-specific tasks in the social colony[82]. By contrast, the ability of the bee to detect toxic substances may have deteriorated owing to the presence of nurturing pre-adult individuals in the hive and a symbiotic relationship with certain plants. It was recently reported that the red flour beetle, *Tribolium castaneum*, has an even larger repertoire of OR genes than the honeybee[83,84]. However, many intact genes in this species do not seem to be expressed, probably because of mutations that affect the regulation of gene expression[83].

The extent of interspecific variation in the CR genes seems to be smaller in insects than in vertebrates. FIGURE 2a shows how the evolution of tetrapod animals (from *Xenopus* spp. to humans) has occurred during the last 360 million years. Similarly, FIG. 2b shows how the insect species considered in this Review have evolved during nearly the same evolutionary time. However, the variation in the number of olfactory or taste receptor genes is much lower in insects than in vertebrates. This is particularly evident in the red flour beetle if only the intact OR genes (146 genes) that are transcribed are considered[83]. The main reason for the difference in the extent of CR gene interspecific variation between vertebrates and insects is probably the greater diversity in ecological conditions, lifestyle, anatomy and physiology in vertebrates compared with insects. Vertebrate species live in water, land and air, and their living conditions are diverse compared with those of the insects considered here, although this is a subjective view.

With regard to gene expression, there is another explanation at least for the difference in the interspecific variation of vertebrate and insect OR genes[50,85]: there are differences between the gene expression regulatory systems of vertebrates and insects (BOX 1). In mice, a group of OR genes exists as a cluster on a chromosomal region, and it has been suggested that any one of these genes

is randomly chosen to be expressed in each olfactory neuron[86,87]. Therefore, the OR gene expression pattern in mammalian species may remain unchanged even when gains or losses of OR genes occur. However, in *D. melanogaster*, each OR gene tends to exist in an isolated state and is expressed in a deterministic manner in a particular olfactory neuron with the ubiquitously expressed *Or83b* gene[29,88,89]. The expression pattern of OR genes is therefore more stringent in *D. melanogaster*, and if gains or losses in the number of OR genes occur this may disrupt the defined expression pattern. Nevertheless, substantial amounts of gene duplication and gene deletion have occurred in insects, as mentioned below. Therefore, the actual OR gene expression patterns in insects may not be as stringent as the above model of gene expression suggests.

Genomic drift and adaptation

The number of chemosensory receptor genes varies extensively among different vertebrate or insect species, owing to repeated gene duplication and deletion, as well as the evolution of pseudogenes. This type of evolutionary change is called birth-and-death evolution[90,91]. Most multigene families are subject to this mode of evolution[91], but OR and VR genes represent extreme cases.

A simple way to detect birth-and-death evolution is to examine the numbers of gains and losses of genes for each branch of the phylogenetic tree of the species studied. These numbers can be estimated using different statistical methods[85,92]. FIGURE 4b shows such estimates for the birth-and-death process of OR genes for eight mammalian species[50]. The estimates show that the number of gains and losses of genes can be as large as several hundreds for every branch of the tree. For example, the opossum lineage gained 759 genes and lost 63 genes from the MRCA of opossums and placental mammals. Even if two extant species have similar numbers of genes, the number of gains and losses of genes can be substantial, as for humans and macaques. This observation indicates that the gene contents of the two species may be different even if the number of genes is nearly the same in both of the species. A similar result has been reported for humans and chimpanzees[44].

Birth-and-death evolution can be caused by both adaptation and random events. Gene duplication occurs primarily by unequal crossover, excluding rare genome-duplication events, and the occurrence of gene duplication is dictated by chance events. However, fixation of duplicate genes in the genome can be aided by natural selection or can occur by random effects. If a higher number of gene copies enhance the ability of an organism to adapt to a particular environment, the number of genes will increase. We have seen that when teleost fishes evolved into land vertebrates there was a large increase in the number of OR genes that detect airborne odorants, such that current mammalian species have hundreds to thousands of OR genes. We have also discussed the possibility that the small number of functional OR genes in aquatic and semi-aquatic mammals is caused by a reduction in gene number during the process of adaptation to new environments.

Nevertheless, the large numbers of gains and losses of genes observed for almost all branches of the mammalian tree suggest that a substantial portion of gene number changes must be due to random gene-duplication and inactivation events. This random change in gene copy number is called genomic drift, analogous to the random genetic drift of gene frequencies in population genetics[93]. The data shown in FIG. 4c suggest that genomic drift also occurs in *Drosophila* spp. OR genes because there is no directional change in the number of genes. However, the extent of genomic drift seems to be lower in *Drosophila* spp. than in mammals, because the number of gains and losses of OR genes are much smaller in *Drosophila* spp. than in mammals and *Drosophila* spp. have a more rigid regulatory system of gene expression, as explained earlier. It should be noted that pseudogenes are also subject to genomic drift, because pseudogenes are non-functional and are thought to evolve in a neutral manner[56]. The inter-order variation of OR pseudogenes in mammals seems to be as large as the variation of OR functional genes.

Copy-number variation within species

If genomic drift occurs frequently, one would expect that a substantial amount of copy-number variation is generated among individuals of the same species. In recent years, many authors[94–96] have studied the extent of copy-number variation in various regions of the human genome (see REFS 97,98 for reviews). These studies have shown that the extent of copy-number variation is at least as great as the extent of allelic polymorphism at single loci. The most extensive study was carried out by Redon *et al.*[96], who studied copy-number variations (CNVs) among 90 African, 90 Asian and 90 European individuals. The original analysis of these data was crude because gene ontology data were used for gene classification, but more detailed analyses were later conducted for each gene family (FIG. 5). Gene-family analysis has shown that CR gene families generally contain a substantial number of CNVs[8,99,100]. The most conspicuous example was the OR gene family, in which approximately 30% of the functional OR genes were polymorphic with respect to copy number among 270 humans[8]. The extent of copy-number variation in CR genes in mice is also substantial but slightly lower than that in humans[99]. Studies of other gene families also showed a substantial number of CNVs[101–103], but the extent of variation was generally smaller than that observed for the CR gene families in humans[8,96,104] and mice[105,106].

Another interesting observation was that there is an almost normal distribution of the number of gene copies per individual when a large number of genes are examined, as for the OR genes[8] (FIG. 5b). This type of normal distribution is expected to occur when the gene family is subject to a random birth-and-death process or when there are a large number of factors (for example, gene fixation by genetic drift, hitch-hiking, gene interaction, and gene loss by deletion or inactivation) that contribute to the increase or decrease in the number of genes. Notably, both the numbers of functional OR genes and pseudogenes have similar standard deviations (FIG. 5b). Because the evolutionary changes that occur in pseudogenes are presumably neutral, these results suggest that the copy

Birth-and-death evolution
An evolutionary mechanism that occurs in multigene families, in which new genes are created by gene duplication and some are retained in the genome for a long time as functional genes, but other genes are inactivated or eliminated from the genome.

Genetic drift
The random change of the allele frequency in populations.

Hitch-hiking
The increase in the frequency of a neutral allele at a locus that is physically linked to an advantageous allele at a different locus.

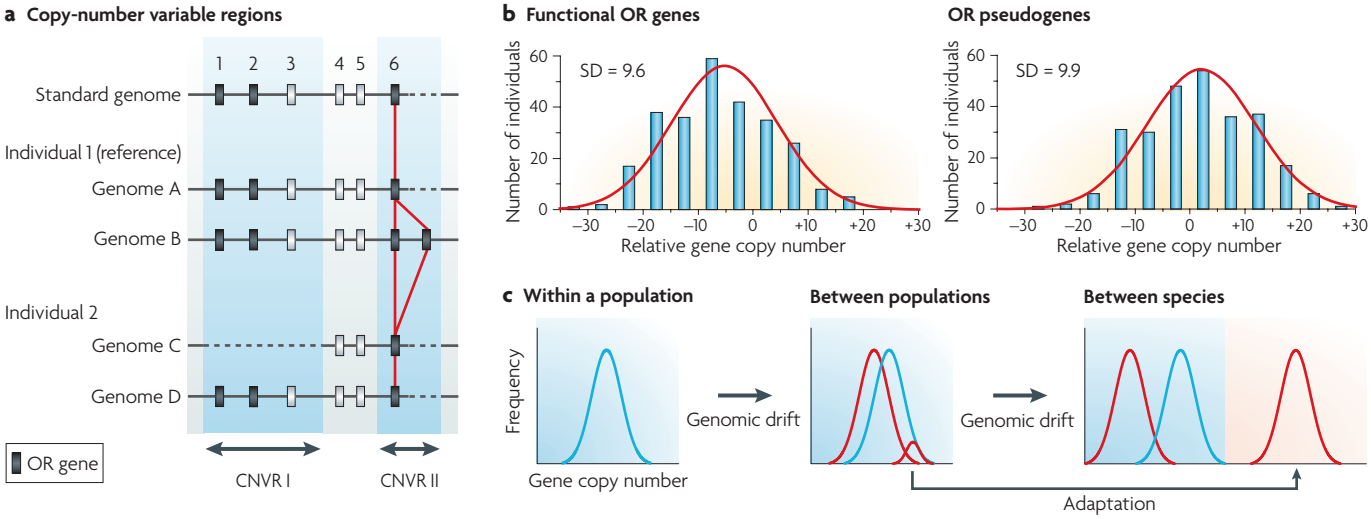


Figure 5 | **Copy-number variation and genomic drift. a** | A simple example showing how copy-number variation is measured. Four genomes (haplotypes) from two diploid individuals are shown. Individual 1, who has the genomes A and B, is arbitrarily chosen as the reference individual. Genome A contains six genetic loci that each have one copy of a gene, whereas genome B has seven gene copies at the six loci because a gene at locus 6 is duplicated compared with the standard genome, for which the whole genome sequence is available. The genes at loci 1–3 are deleted from genome C of individual 2, and the genes contained in genome D are the same as those present in genome A. In this case, there are four copy-number polymorphic loci in copy-number variable regions (CNVRs) I and II and three of these are for olfactory receptor (OR) genes (grey boxes). Because the copy numbers in two genomes of an individual cannot be measured separately, we consider the total number of gene copies in each individual. The total numbers of OR genes in individuals 1 and 2 are seven and four, respectively. Therefore, the number of OR genes for individual 2 relative to the number for the reference individual (individual 1) is calculated by 4 – 7 = –3. The relative number for the reference individual is defined as 0. **b** | Distributions of relative copy number of OR genes in humans. The relative copy number represents the difference in copy number between a sampled individual and the reference individual, as shown in part **a**. The curve represents the normal distribution. **c** | Genomic drift is a random process of copy-number changes that occur by duplication, deletion and inactivation of genes[8,93]. In this case, a distribution of gene copy number will follow the normal distribution if the number of gene copies is sufficiently large. Consequently, a natural population has a substantial amount of copy-number variation (shown by the curve on the left) by genomic drift as long as the copy number is within a range determined by functional requirements (shown by the background colour). In the case of chemosensory receptor genes, this copy-number range is generally large. Therefore, when a population is separated into two geographic populations, these populations can have different distributions of copy number (middle graph). When these populations evolve into different species, the copy-number difference may be even larger owing to genomic drift (shown by the graph on the right). A new species can also be generated when a group of individuals who have a larger number of genes (the small peak in the middle diagram) moves to a new niche where a larger number of genes is more advantageous (the right peak on the right graph). SD, standard deviation. Parts **b** and **c** are modified, with permission, from REF. 8 © (2007) National Academy of Sciences.

number of functional OR genes is also determined largely by random factors[8]. Young and colleagues[100] also reported that CNVs of OR genes accumulate in a neutral way in humans. These observations suggest that although the large variation in gene copy number among distantly related species is apparently caused by adaptation, the intraspecific variation is almost neutral. In other words, the number of gene copies may vary considerably as long as this number is within the upper and lower limits determined by the physiological requirements of the organism, and the extent of the copy-number variation determined in this way is considerable.

Functional changes and the role of selection

So far, we have discussed the evolution of chemosensation in terms of gene copy number and related the long-term evolution of chemosensation to the genetic change within populations. This type of study has only become possible because genome sequence data are now available for many different species. In classic evolutionary genetics, it is customary to identify polymorphic alleles and examine the allele frequency change caused by natural selection or genetic drift. Theoretically, this approach can also be used to study CR genes, but this may be challenging because there are many CR gene copies, with the exception of T1R genes, and it is often difficult to detect clear relationships between genotypes and phenotypes.

Recently, several groups have used another approach to study the effect of natural selection at the DNA sequence level: they estimated the d_N:d_S ratio for a set of homologous genes or specific set of codons. When a large number of full-length sequences were examined using non-likelihood methods[43], positive Darwinian selection was suggested to have occurred in only a small number of cases[57,107,108]. However, when a likelihood method with a codon substitution model in the computer program PAML[109] was used,

signatures of positive selection were often obtained[110–114]. These results indicated that CR genes, particularly OR genes, may have evolved under positive selection.

However, it is necessary to interpret these results with caution. First, non-synonymous nucleotide substitutions measured by d_N do not necessarily indicate positive selection, because most amino-acid substitutions seem to be almost neutral, as shown by early molecular evolutionists[115–117]. In a recent experimental study of the evolution of visual (rhodopsin) genes from a large number of vertebrate species, about 94% of a total of 191 amino-acid substitutions were shown to be roughly neutral and only 15 substitutions resulted in functional changes in the visual genes[118]. Second, the likelihood method of inferring positive selection is known to frequently yield false-positive results[119–122]. When Yokoyama *et al.*[118] used this method, none of the amino-acid sites that were statistically predicted to be under selection matched with the adaptive sites that have been demonstrated experimentally. These authors therefore emphasized the importance of experimental verification of statistical inference using site-directed mutagenesis. Third, the relationship between amino-acid substitution and change in fitness is complex, particularly for a character controlled by a large number of genes. In the case of OR genes, it is sometimes possible to identify the effect of a particular mutant allele on olfactory ability and evaluate the evolutionary significance of this effect by specific experimental methods[77,123]. However, because olfaction occurs with a combinatory set of OR genes, the study of a single mutation may not reveal the adaptive importance of that particular mutation. Furthermore, the fitness or the expected number of offspring per individual may not be severely affected by a set of polymorphic alleles, because fitness is determined by many other characteristics in addition to olfaction, such as vision, physical strength, mental ability and resistance to diseases. Even individuals with no sense of smell (anosmics) do not seem to have any fertility problems. Therefore, it will be a challenge to study the evolution of chemosensation by natural selection, except by using a gene family with a small number of genes.

Three eminent evolutionists in the twentieth century, R. A. Fisher, E. B. Ford and J. Huxley, famously proposed that the polymorphism of the *T* and *t* alleles at the locus of the phenylthiocarbamide (PTC; bitter) taste receptor, TAS2R38, in humans is caused by overdominant selection, because the polymorphism is shared by both humans and chimpanzees and therefore must have been maintained in the population for a long time[124]. A recent study of the molecular basis of the PTC polymorphism has shown that the *T* and *t* alleles in humans and chimpanzees are different and originated independently in recent years, and that the *t* allele in the chimpanzee contains an interrupted reading frame[125]. These results do not support the hypothesis of Fisher *et al.*, but from examining the pattern of DNA polymorphism at this locus, it is still argued that the PTC polymorphism in humans is due to overdominant selection.

A similar study that used the population genetics methods described by Tajima[126] and McDonald and Kreitman[127] analysed human and chimpanzee OR genes[57], and some signatures of positive selection were obtained for a number of genes[57]. However, these population genetics approaches only provide suggestive evidence[121,128], and experimental studies are required to obtain a definitive conclusion. However, at present we are left with only suggestive evidence, because it is difficult to experimentally verify natural selection. This challenge is partly due to the large number of stochastic errors associated with allele frequency changes within populations. This problem can be largely overcome by comparing distantly related species to study long-term evolution, as discussed above.

Conclusions

We have described how the evolution of CR genes occurs by both directional selection and random forces, as for other characteristics. However for CR genes, a large number of genes is involved, with the exception of the T1R and TAAR genes, and this number often changes dramatically during the evolutionary process. The basic processes involved in the evolution of CR genes are gene duplication and deletion, which occur almost randomly. We have called this random contribution to changes in gene number genomic drift. Genomic drift is a new aspect of CR gene evolution that had been previously unrecognized[93]. It has a dual role in evolution and causes both adaptive and neutral changes in phenotypic characteristics. The CNVs observed within populations may be largely neutral, but if a population moves to a new niche, a portion of these may be used selectively for that population to adapt to the new niche (FIG. 5c). In this case, a species-specific expansion of a certain group of CR genes may occur by further gene duplication.

We have described several cases in which gene expansion occurred without a clear adaptive explanation (for example, the opossum OR genes). In these cases, gene expansion may have occurred by genomic drift. In fact, if we introduce the concept of genomic drift, many such observations can be explained. However, this concept should not be abused to produce various just-so stories (ad hoc hypotheses). It is important to study the number of genes in closely related and distantly related species to identify the adaptive and random forces of evolution. It is also important to consider other aspects of evolution and examine whether adaptive forces are involved, as we examined the *w* value in relation to the vision-priority hypothesis. Another problem that complicates the evolutionary study of CR genes is the presence of a large number of pseudogenes. The number of OR pseudogenes is particularly large in cows and *Xenopus* spp. Why are so many pseudogenes retained in the genome? In this Review, we have considered the possibility that pseudogenes are transcribed and that the transcripts regulate the expression patterns of functional genes. There is currently no evidence to support this proposition, but it is an interesting possibility for further investigation.

According to the widely accepted neo-Darwinian theory of evolution, the evolution of phenotypic characters is almost exclusively determined by natural selection and mutation merely provides raw genetic material for natural selection to produce novel characteristics[129–131].

Overdominant selection
A form of selection caused by heterozygote advantage.

This view has recently been challenged by Nei[93,132], Stoltzfus[133], Lynch[134] and others, who argue that mutation is the driving force of evolution and that natural selection is of secondary importance (neomutationism). In the context of this Review, mutation refers to gene duplication or gene inactivation, whereas natural selection is implied by the occurrence of adaptation to a particular environment[93]. Although we cannot measure the relative importance of adaptive and neutral evolution, data from the evolution of chemosensation clearly show that the neutral evolution of phenotypic characters exists and is an important process.

1. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. The complete human olfactory subgenome. *Genome Res.* **11**, 685–702 (2001).
2. Zozulya, S., Echeverri, F. & Nguyen, T. The human olfactory receptor repertoire. *Genome Biol.* **2**, 0018.1–0018.12 (2001).
In References 1 and 2, a bioinformatics approach was used for the first time to identify all of the OR genes in the human genome.
3. Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci.* **5**, 124–133 (2002).
4. Young, J. M. *et al.* Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum. Mol. Genet.* **11**, 535–546 (2002).
5. Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D. & Paabo, S. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* **2**, e5 (2004).
6. Niimura, Y. & Nei, M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc. Natl Acad. Sci. USA* **102**, 6039–6044 (2005).
7. Liman, E. R. Use it or lose it: molecular evolution of sensory signaling in primates. *Pflugers Arch.* **453**, 125–131 (2006).
8. Nozawa, M., Kawahara, Y. & Nei, M. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc. Natl Acad. Sci. USA* **104**, 20421–20426 (2007).
This study empirically showed that genomic drift is an important mechanism in the evolution of chemosensory receptor genes.
9. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
This is a seminal paper that identified OR genes for the first time and estimated that there are approximately 1,000 OR genes in the mammalian genome.
10. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645–650 (2006).
11. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
12. Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**, 763–773 (1997).
13. Matsunami, H. & Buck, L. B. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**, 775–784 (1997).
14. Li, X. *et al.* Human receptors for sweet and umami taste. *Proc. Natl Acad. Sci. USA* **99**, 4692–4696 (2002).
15. Adler, E. *et al.* A novel family of mammalian taste receptors. *Cell* **100**, 693–702 (2000).
16. Matsunami, H., Montmayeur, J. P. & Buck, L. B. A family of candidate taste receptors in human and mouse. *Nature* **404**, 601–604 (2000).
17. Feldmesser, E. *et al.* Widespread ectopic expression of olfactory receptor genes. *BMC Genomics* **7**, 121 (2006).
18. Borowsky, B. *et al.* Trace amines: identification of a family of mammalian G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **98**, 8966–8971 (2001).
19. Zucchi, R., Chiellini, G., Scanlan, T. S. & Grandy, D. K. Trace amine-associated receptors and their ligands. *Br. J. Pharmacol.* **149**, 967–978 (2006).
20. Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
21. Brennan, P. A. & Zufall, F. Pheromonal communication in vertebrates. *Nature* **444**, 308–315 (2006).
22. Baxi, K. N., Dorries, K. M. & Eisthen, H. L. Is the vomeronasal system really specialized for detecting pheromones? *Trends Neurosci.* **29**, 1–7 (2006).
23. Ikeda, K. New seasonings. *Chem. Senses* **27**, 847–849 (2002).
This is an English translation of Ikeda's original paper in Japanese from 1909 on the discovery of the umami taste.
24. Chandrashekar, J., Hoon, M. A., Ryba, N. J. & Zuker, C. S. The receptors and cells for mammalian taste. *Nature* **444**, 288–294 (2006).
25. Bachmanov, A. A. & Beauchamp, G. K. Taste receptor genes. *Annu. Rev. Nutr.* **27**, 389–414 (2007).
26. Clyne, P. J. *et al.* A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327–338 (1999).
27. Gao, Q. & Chess, A. Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* **60**, 31–39 (1999).
28. Clyne, P. J., Warr, C. G. & Carlson, J. R. Candidate taste receptors in *Drosophila*. *Science* **287**, 1830–1834 (2000).
29. Bargmann, C. I. Comparative chemosensation from receptors to ecology. *Nature* **444**, 295–301 (2006).
30. Vosshall, L. B. & Stocker, R. F. Molecular architecture of smell and taste in *Drosophila*. *Annu. Rev. Neurosci.* **30**, 505–533 (2007).
31. Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100**, 14537–14542 (2003).
This is the first comprehensive study about the evolution of the chemosensory receptor gene superfamily in *Drosophila* spp.
32. Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86–90 (2007).
33. Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019–1029 (2003).
34. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
35. Larsson, M. C. *et al. Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).
36. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
37. Godfrey, P. A., Malnic, B. & Buck, L. B. The mouse olfactory receptor gene family. *Proc. Natl Acad. Sci. USA* **101**, 2156–2161 (2004).
38. Malnic, B., Godfrey, P. A. & Buck, L. B. The human olfactory receptor gene family. *Proc. Natl Acad. Sci. USA* **101**, 2584–2589 (2004).
39. Niimura, Y. & Nei, M. Evolution of olfactory receptor genes in the human genome. *Proc. Natl Acad. Sci. USA* **100**, 12235–12240 (2003).
40. Berghard, A. & Dryer, L. A novel family of ancient vertebrate odorant receptors. *J. Neurobiol.* **37**, 383–392 (1998).
41. Freitag, J., Beck, A., Ludwig, G., von Buchholtz, L. & Breer, H. On the origin of the olfactory receptor family: receptor genes of the jawless fish (*Lampetra fluviatilis*). *Gene* **226**, 165–174 (1999).
42. Niimura, Y. & Nei, M. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene* **346**, 13–21 (2005).
43. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, 2000).
44. Go, Y. & Niimura, Y. Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. *Mol. Biol. Evol.* **25**, 1897–1907 (2008).
45. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
46. Bowmaker, J. K. in *Evolution of the Eye and Visual System* (eds Cronly-Dillon, J. R. & Gregory, R. L.) 63–81 (CRC, Boca Raton, 1991).
47. Steiger, S. S., Fidler, A. E., Valcu, M. & Kempenaers, B. Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? *Proc. Biol. Sci.* **275**, 2309–2317 (2008).
48. Laska, M., Seibt, A. & Weber, A. 'Microsmatic' primates revisited: olfactory sensitivity in the squirrel monkey. *Chem. Senses* **25**, 47–53 (2000).
49. Shephard, G. M. The human sense of smell: are we better than we think? *PLoS Biol.* **2**, 572–575 (2004).
50. Niimura, Y. & Nei, M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* **2**, e708 (2007).
This paper showed that large numbers of gene gains and losses occur during the evolution of mammalian OR genes.
51. Pettigrew, J. D. Electroreception in monotremes. *J. Exp. Biol.* **202**, 1447–1454 (1999).
52. Oelschlager, H. H. & Kemp, B. Ontogenesis of the sperm whale brain. *J. Comp. Neurol.* **399**, 210–228 (1998).
53. Freitag, J., Ludwig, G., Andreini, I., Rossler, P. & Breer, H. Olfactory receptors in aquatic and terrestrial vertebrates. *J. Comp. Physiol. A* **183**, 635–650 (1998).
54. Kishida, T., Kubota, S., Shirayama, Y. & Fukami, H. The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biol. Lett.* **3**, 428–430 (2007).
55. McGowen, M. R., Clark, C. & Gatesy, J. The vestigial olfactory receptor subgenome of odontocete whales: phylogenetic congruence between gene-tree reconciliation and supermatrix methods. *Syst. Biol.* **57**, 574–590 (2008).
56. Li, W. H., Gojobori, T. & Nei, M. Pseudogenes as a paradigm of neutral evolution. *Nature* **292**, 237–239 (1981).
57. Gilad, Y., Bustamante, C. D., Lancet, D. & Paabo, S. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* **73**, 489–501 (2003).
58. Zhang, X. *et al.* Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biol.* **8**, R86 (2007).
59. Balakirev, E. S. & Ayala, F. J. Pseudogenes: are they "junk" or functional DNA? *Annu. Rev. Genet.* **37**, 123–151 (2003).
60. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653–1655 (2006).
61. Zheng, D. & Gerstein, M. B. The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends Genet.* **23**, 219–224 (2007).
62. Lai, P. C. *et al.* An olfactory receptor pseudogene whose function emerged in humans. *Nature Precedings* 2 Nov 2007 (doi:10.1038/npre.2007.1290.1).
63. Grus, W. E. & Zhang, J. Distinct evolutionary patterns between chemoreceptors of 2 vertebrate olfactory systems and the differential tuning hypothesis. *Mol. Biol. Evol.* **25**, 1593–1601 (2008).
64. Gloriam, D. E. *et al.* The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish. *Mol. Phylogenet. Evol.* **35**, 470–482 (2005).
65. Hashiguchi, Y. & Nishida, M. 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**, 2099–2107 (2007).
66. Grus, W. E., Shi, P., Zhang, Y. P. & Zhang, J. Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals. *Proc. Natl Acad. Sci. USA* **102**, 5767–5772 (2005).
67. Young, J. M., Kambere, M., Trask, B. J. & Lane, R. P. Divergent V1R repertoires in five species: amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs. *Genome Res.* **15**, 231–240 (2005).

68. Grus, W. E., Shi, P. & Zhang, J. Largest vertebrate vomeronasal type 1 receptor gene repertoire in the semiaquatic platypus. *Mol. Biol. Evol.* **24**, 2153–2157 (2007).
69. Young, J. M. & Trask, B. J. V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends Genet.* **23**, 212–215 (2007).
70. Liman, E. R. & Innan, H. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc. Natl Acad. Sci. USA* **100**, 3328–3332 (2003).
71. Zhang, J. & Webb, D. M. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl Acad. Sci. USA* **100**, 8337–8341 (2003).
72. Rodriguez, I., Greer, C. A., Mok, M. Y. & Mombaerts, P. A putative pheromone receptor gene expressed in human olfactory mucosa. *Nature Genet.* **26**, 18–19 (2000).
73. Shirokova, E., Raguse, J. D., Meyerhof, W. & Krautwurst, D. The human vomeronasal type-1 receptor family — detection of volatiles and cAMP signaling in HeLa/Olf cells. *FASEB J.* **22**, 1416–1425 (2008).
74. Shepherd, G. M. Smells, brains and hormones. *Nature* **439**, 149–151 (2006).
75. Shi, P. & Zhang, J. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res.* **17**, 166–174 (2007).
76. Bonadonna, F., Miguel, E., Grosbois, V., Jouventin, P. & Bessiere, J. M. Individual odor recognition in birds: an endogenous olfactory signature on petrels' feathers? *J. Chem. Ecol.* **33**, 1819–1829 (2007).
77. Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B. & Matsunami, H. Genetic variation in a human odorant receptor alters odour perception. *Nature* **449**, 468–472 (2007).
78. Shi, P. & Zhang, J. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol. Biol. Evol.* **23**, 292–300 (2006).
79. Li, X. *et al.* Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. *PLoS Genet.* **1**, 27–35 (2005).
80. Hashiguchi, Y., Furuta, Y., Kawahara, R. & Nishida, M. Diversification and adaptive evolution of putative sweet taste receptors in threespine stickleback. *Gene* **396**, 170–179 (2007).
81. McBride, C. S., & Arguello, J. R. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395–1416 (2007).
82. Robertson, H. M. & Wanner, K. W. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res.* **16**, 1395–1403 (2006).
83. Engsontia, P. *et al.* The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* **38**, 387–397 (2008).
84. Abdel-Latief, M. A family of chemoreceptors in *Tribolium castaneum* (Tenebrionidae: Coleoptera). *PLoS ONE* **2**, e1319 (2007).
85. Nozawa, M. & Nei, M. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc. Natl Acad. Sci. USA* **104**, 7122–7127 (2007).
86. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor–one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
87. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
88. Ray, A., van Naters, W. G., Shiraiwa, T. & Carlson, J. R. Mechanisms of odor receptor gene choice in *Drosophila*. *Neuron* **53**, 353–369 (2007).
89. Ray, A., van der Goes van Naters, W. & Carlson, J. R. A regulatory code for neuron-specific odor receptor expression. *PLoS Biol.* **6**, e125 (2008).
90. Nei, M., Gu, X. & Sitnikova, T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl Acad. Sci. USA* **94**, 7799–7806 (1997).
91. Nei, M. & Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**, 121–152 (2005).
92. Nam, J. & Nei, M. Evolutionary change of the numbers of homeobox genes in bilateral animals. *Mol. Biol. Evol.* **22**, 2386–2394 (2005).
93. Nei, M. The new mutation theory of phenotypic evolution. *Proc. Natl Acad. Sci. USA* **104**, 12235–12242 (2007).
94. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
95. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
96. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
This study showed that the extent of copy number variation in the human genome is approximately 12% and indicates that such variation may be more important than single nucleotide polymorphisms in the generation of genetic variation.
97. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7**, 85–97 (2006).
98. Cooper, G. M., Nickerson, D. A. & Eichler, E. E. Mutational and selective effects on copy-number variants in the human genome. *Nature Genet.* **39**, S22–S29 (2007).
99. Nozawa, M. & Nei, M. Genomic drift and copy number variation of chemosensory receptor genes in humans and mice. *Cytogenet. Genome Res.* (in the press).
100. Young, J. M. *et al.* Extensive copy-number variation of the human olfactory receptor gene family. *Am. J. Hum. Genet.* **83**, 228–242 (2008).
101. Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet.* **1**, 129–139 (2005).
102. Fanciulli, M. *et al. FCGR3B* copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature Genet.* **39**, 721–723 (2007).
103. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39**, 1256–1260 (2007).
104. Nguyen, D. Q., Webber, C. & Ponting, C. P. Bias of selection on human copy-number variants. *PLoS Genet.* **2**, e20 (2006).
105. Cutler, G., Marshall, L. A., Chin, N., Baribault, H. & Kassner, P. D. Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.* **17**, 1743–1754 (2007).
106. Graubert, T. A. *et al.* A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* **3**, e3 (2007).
107. Kondo, R., Kaneko, S., Sun, H., Sakaizumi, M. & Chigusa, S. I. Diversification of olfactory receptor genes in the Japanese medaka fish, *Oryzias latipes*. *Gene* **282**, 113–120 (2002).
108. Gimelbrant, A. A., Skaletsky, H. & Chess, A. Selective pressures on the olfactory receptor repertoire since the human–chimpanzee divergence. *Proc. Natl Acad. Sci. USA* **101**, 9019–9022 (2004).
109. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
110. Clark, A. G. *et al.* Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
111. Emes, R. D., Beatson, S. A., Ponting, C. P. & Goodstadt, L. Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. *Genome Res.* **14**, 591–602 (2004).
112. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
113. Shi, P., Bielawski, J. P., Yang, H. & Zhang, Y. P. Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J. Mol. Evol.* **60**, 566–576 (2005).
114. Tunstall, N. E., Sirey, T., Newcomb, R. D. & Warr, C. G. Selective pressures on *Drosophila* chemosensory receptor genes. *J. Mol. Evol.* **64**, 628–636 (2007).
115. Zuckerkandl, E. & Pauling, L. in *Evolving Genes and Proteins* (eds Bryson, V. & Vegel, H. J.) 97–166 (Academic, New York, 1965).
116. King, J. L. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
117. Perutz, M. F. Species adaptation in a protein molecule. *Mol. Biol. Evol.* **1**, 1–28 (1983).
118. Yokoyama, S., Tada, T., Zhang, H. & Britt, L. Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl Acad. Sci. USA* **105**, 13480–13485 (2008).
119. Suzuki, Y. & Nei, M. False-positive selection identified by ML-based methods: examples from the *Sig1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* **21**, 914–921 (2004).
120. Hughes, A. L. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* **99**, 364–373 (2007).
121. Hughes, A. L. The origin of adaptive phenotypes. *Proc. Natl Acad. Sci. USA* **105**, 13193–13194 (2008).
122. Suzuki, Y. False-positive results obtained from the branch-site test of positive selection. *Genes Genet. Syst.* **83**, 331–338 (2008).
123. Menashe, I. *et al.* Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol.* **5**, e284 (2007).
This study showed that copy-number polymorphism of OR genes affects odorant sensitivity in different individuals.
124. Fisher, R. A., Ford, E. B. & Huxley, J. Taste-testing the anthropoid apes. *Nature* **144**, 750 (1939).
125. Wooding, S. *et al.* Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature* **440**, 930–934 (2006).
126. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
127. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
128. Nei, M. Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* **22**, 2318–2342 (2005).
129. Mayr, E. *Animal Species and Evolution* (Belknap Press of Harvard Univ., Massachusetts, 1963).
130. Futuyma, D. J. *Evolution* (Sinauer Associates, Massachusetts, 2005).
131. Dobzhansky, T. *Genetics and the Origin of Species*. 3rd edn (Columbia Univ. Press, New York, 1951).
132. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
133. Stoltzfus, A. Mutationism and the dual causation of evolutionary change. *Evol. Dev.* **8**, 304–317 (2006).
134. Lynch, M. *The Origins of Genome Architecture* (Sinauer Associates, Massachusetts, 2007).
135. Lindemann, L. *et al.* Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics* **85**, 372–385 (2005).
136. Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C. & Ritchie, M. G. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol. Ecol.* **17**, 1648–1657 (2008).
137. Hill, C. A. *et al.* G protein-coupled receptors in *Anopheles gambiae*. *Science* **298**, 176–178 (2002).
138. Kent, L. B., Walden, K. K. & Robertson, H. M. The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem. Senses* **33**, 79–93 (2008).
139. Bohbot, J. *et al.* Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Mol. Biol.***16**, 525–537 (2007).
140. Wanner, K. W. *et al.* Female-biased expression of odorant receptor genes in the adult antennae of the silkworm, *Bombyx mori*. *Insect Mol. Biol.* **16**, 107–119 (2007).

Acknowledgements
We thank Y. Go for providing unpublished data. This work was supported by National Institutes of Health grant GM020293 to M.N. and Japanese Government Grants 17710162 and 20770192 to Y.N.

DATABASES
Entrez Gene:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
OR1E3 | *T1R1* | *T1R2* | *T1R3*
UniProtKB: http://www.uniprot.org
Or83b | TAS2R38

FURTHER INFORMATION
Masatoshi Nei's homepage:
http://www.bio.psu.edu/People/Faculty/Nei/Lab/
Yoshihito Niimura's homepage:
http://bioinfo.tmd.ac.jp/~niimura/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF