

Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds?

Silke S. Steiger^{1,*}, Andrew E. Fidler², Mihai Valcu¹ and Bart Kempnaers¹

¹Department of Behavioural Ecology and Evolutionary Genetics, Max-Planck Institute for Ornithology, PO Box 1564, 82319 Starnberg, Germany

²Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand

Among vertebrates, the sense of smell is mediated by olfactory receptors (ORs) expressed in sensory neurons within the olfactory epithelium. Comparative genomic studies suggest that the olfactory acuity of mammalian species correlates positively with both the total number and the proportion of functional OR genes encoded in their genomes. In contrast to mammals, avian olfaction is poorly understood, with birds widely regarded as relying primarily on visual and auditory inputs. Here, we show that in nine bird species from seven orders (blue tit, *Cyanistes caeruleus*; black coucal, *Centropus grillii*; brown kiwi, *Apteryx australis*; canary, *Serinus canaria*; galah, *Eolophus roseicapillus*; red jungle fowl, *Gallus gallus*; kakapo, *Strigops habroptilus*; mallard, *Anas platyrhynchos*; snow pterel, *Pagodroma nivea*), the majority of amplified OR sequences are predicted to be from potentially functional genes. This finding is somewhat surprising as one previous report suggested that the majority of OR genes in an avian (red jungle fowl) genomic sequence are non-functional pseudogenes. We also show that it is not the estimated proportion of potentially functional OR genes, but rather the estimated total number of OR genes that correlates positively with relative olfactory bulb size, an anatomical correlate of olfactory capability. We further demonstrate that all the nine bird genomes examined encode OR genes belonging to a large gene clade, termed γ -c, the expansion of which appears to be a shared characteristic of class Aves. In summary, our findings suggest that olfaction in birds may be a more important sense than generally believed.

Keywords: olfaction; bird; olfactory bulb; pseudogene; nocturnality; ecological adaptation

1. INTRODUCTION

Olfactory receptors (ORs) expressed in sensory neurons within the olfactory epithelium constitute the molecular basis of the sense of smell among vertebrates (Buck & Axel 1991; Gaillard *et al.* 2003). OR genes are small (approx. 1000 bp), intronless (Young & Trask 2002; Mombaerts 2004) and are thought to evolve rapidly, following a 'birth-and-death' model (Nei *et al.* 1997). Both the size of the OR gene family and the proportion of OR genes that are non-functional (i.e. pseudogenes) vary widely between vertebrate genomes (size range: 100–2130 in pufferfish, *Fugu rubripes*, and cow, *Bos taurus*, respectively; predicted functional proportion range: 40–80% in human and mouse, respectively; Mombaerts 2004; Niimura & Nei 2006, 2007). Comparative genomic studies suggest that the olfactory acuity of mammalian species correlates positively with both the total number and the proportion of functional OR genes encoded in their genomes (Rouquier *et al.* 2000; Gilad *et al.* 2004; Niimura & Nei 2006, 2007). The total number of OR genes in a genome may reflect how many different scents can be detected and distinguished (Niimura & Nei 2006). The proportion of functional OR genes provides insights into the selective pressures that have acted on the OR genes

(Rouquier *et al.* 2000; Niimura & Nei 2006). For example, if olfaction has become less important during the evolutionary history of a species, an associated relaxation of conservative selection pressure may have led to an increase in the number of pseudogenes (i.e. no selection against loss-of-function mutations). Indeed, it has been suggested that a decline in the proportion of functional OR genes in the human genome is associated with a less keen sense of smell when compared with other primates (Rouquier *et al.* 2000; Gilad *et al.* 2004). OR genes have been studied extensively in fishes and mammals (Niimura & Nei 2006). By contrast, far less is known about avian OR genes. This may reflect the general belief that birds lack a well-developed sense of smell, although behavioural studies have shown that some bird species use their sense of smell to navigate (Papi 1991), forage (Wenzel 1968; Nevitt *et al.* 2008) or distinguish individuals (Bonadonna & Nevitt 2004; for reviews, see Roper 1999; Hagelin 2006; Hagelin & Jones 2007).

To date, avian OR gene sequence data have been limited to the domestic chicken (*Gallus gallus domesticus*) and its wild progenitor, the red jungle fowl (*Gallus gallus*; Leibovici *et al.* 1996; Nef *et al.* 1996; International Chicken Genome Sequencing Consortium 2004; Niimura & Nei 2005; Lagerström *et al.* 2006; but see Eriksson *et al.* 2008). An analysis of a draft (BUILD v. 1.1) *G. gallus* genomic sequence reported that (i) the OR gene repertoire consists of approximately 550 members, (ii) the predicted

* Author for correspondence (steiger@orn.mpg.de).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2008.0607> or via <http://journals.royalsociety.org>.

proportion of potentially functional OR genes was approximately 15%, and (iii) the majority of the *G. gallus* OR genes clustered within a single large clade, denoted group- γ -c (Niimura & Nei 2005). The group- γ -c clade appears to have expanded in size after the separation of the avian and mammalian lineages (Niimura & Nei 2005) and represents an expansion of OR genes similar to the human OR5U1 and OR5BF1 genes (International Chicken Genome Sequencing Consortium 2004; Lagerström *et al.* 2006). Note that, because the *G. gallus* genomic sequence analysed (BUILD v. 1.1) was of draft status, the estimated number and proportion of potentially functional OR genes should be considered as underestimates (Niimura & Nei 2005). Indeed, other studies estimated the potentially functional OR gene repertoire of the BUILD v. 1.1 draft *G. gallus* genomic sequence to be either 229 (Lagerström *et al.* 2006) or 283 (International Chicken Genome Sequencing Consortium 2004). The surprisingly large difference in the estimated number of potentially functional OR genes identified in those studies may be attributed to the different bioinformatics search strategies used.

In this study, we estimated the proportion of potentially functional OR genes encoded within the *G. gallus* genome and within the following eight other, taxonomically diverse, bird genomes: the blue tit (*Cyanistes caeruleus*); the black coucal (*Centropus grillii*); the brown kiwi (*Apteryx australis*); the canary (*Serinus canaria*); the galah (*Eolophus roseicapillus*); the kakapo (*Strigops habroptilus*); the mallard (*Anas platyrhynchos*); and the snow petrel (*Pagodroma nivea*). We further investigated whether either the proportion of potentially functional OR genes or the estimated total number of OR genes correlates with the olfactory bulb ratio (OBR), a possible anatomical correlate of olfactory capability (Edinger 1908). OBRs vary widely among avian species (Bang & Cobb 1968) and the nine species we investigated cover the entire range. Additionally, we estimated the total number of OR genes, both potentially functional and non-functional, in the nine species using a sample-coverage approach (Chao & Lee 1992). Finally, we derived phylogenetic trees from predicted OR protein sequences to test whether the recently expanded group- γ -c OR genes are specific to the red jungle fowl or are a shared characteristic of bird genomes.

2. MATERIAL AND METHODS

(a) Amplification and sequencing of OR genes

Blood samples were suspended in Queen's lysis buffer and stored at ambient temperature. Genomic DNA was isolated using a commercial kit (DNeasy tissue kit; Qiagen, Hilden, Germany) and approximately 100 ng was used as a template in subsequent amplification reactions. In total, 10 primers were designed to anneal to evolutionarily conserved coding sequences corresponding to the transmembrane (TM) domain 3 (forward primers) and TM7 (reverse primers) regions of the OR proteins. PCR primer pairs falling into two categories targeting either (i) the non- γ -c OR clade sequences or (ii) the γ -c OR clade sequences were used. To amplify non- γ -c OR sequences, three previously reported forward primers corresponding to the conserved TM3 amino acid sequences of (A)MAYDRY (5'-ATG GCI TAY GAY MGI TA-3' and 5'-GCI ATG GCI TAY GAY MGI TA-3'; Nef *et al.* 1996; Freitag *et al.* 1999) and MAYDRY(V/L)AIC (5'-ATG GCI

TAY GAY MGI TAY STI GCI ATY TG-3'; Leibovici *et al.* 1996) were paired with three reverse primers corresponding to the conserved TM7 amino acid sequences PMLNPLIY (5'-TA DAT IAG IGG RTT IAG CAT IGG-3'), NPFIYS (F/L) (5'-AR ISW RTA DAT RAA IGG RTT-3'; Freitag *et al.* 1999) and PM(L/F)NP (5'-GG RTT IAR CAT IGG-3'; Nef *et al.* 1996). Amplifications were conducted using each forward primer in combination with each reverse primer, thereby generating nine different PCR products. For the amplification of γ -c OR clade sequences, three forward primers corresponding to sequences found to be conserved among the reported red jungle fowl γ -c OR TM3 amino acid sequences ICKPLHY (5'-ATC TGY AAR CCI YTI CAY TA-3') and VAICKPLHY (5'-ATC TGY AAR CCI YTI CAY TA-3' and 5'-RTT GCI ATY TGY AAR CCY CTR CAC TA-3') were used in combination with the reverse primer designed to the conserved TM7 amino acid OR sequence NPFIYS(F/L) (5'-AR ISW RTA DAT RAA IGG RTT-3'; Freitag *et al.* 1999).

All primer pairs were predicted to generate products of approximately 0.5 kb, which represents approximately half of the expected full OR-coding sequence. The PCR was carried out in a final volume of 50.0 μ l, containing 2.0 mM Mg²⁺, dNTPs (0.1 mM), primers (0.8 μ M), 1 U *Taq* DNA polymerase (MBI Fermentas, St Leon-Rot, Germany) and genomic DNA (100 ng) template with thermocycling parameters: 94°C/2 min; 94°C/30 s, 37°C/30 s, ramping from 37 to 72°C at 0.2°C s⁻¹, 72°C/60 s, 5 cycles; 94°C/30 s, 45°C/30 s, 72°C/60 s, 30 cycles; 72°C/7 min; and 4°C/hold.

Amplification products were separated through 2% agarose gels (Nusieve GTG agarose, BioWhittaker Molecular Applications, Rockland, ME, USA), bands were excised and purified (QIAquick Gel Extraction kit, Qiagen) before ligation into a T-tailed cloning vector (pGemT-easy, Promega, Madison, WI, USA). Note that PCR products were not pooled for ligation and transformation. Plasmids having inserts were purified from transformed DH5 α colonies by alkaline lysis (High Pure Plasmid Isolation kit; Roche Diagnostics, Mannheim, Germany) and sequenced by external contractors (MWG Biotech, Ebersberg, Germany; AGOWA, Berlin, Germany).

(b) Sequence analysis and phylogenetic tree construction

We obtained on average 150 \pm 11 s.e.m. (range 98–206) sequences per species. Electropherograms were visually inspected, edited and low-quality sequences discarded. PCR primer sequences were deleted and sequences sharing more than or equal to 98.5% identity, determined using the 'SEQUENCE IDENTITY MATRIX' function of BioEDIT (Hall 1999; <http://www.mbio.ncsu.edu/bioedit/bioedit.html>), were considered to be amplified from a single OR gene (Fuchs *et al.* 2001). This procedure was used to accommodate errors introduced by the amplification itself. It may contribute to an underestimation of the total OR gene number due to the clustering of highly similar, but distinct, paralogues. To confirm that the sequences were partial OR-coding sequences, each sequence was used as a query in a BLAST search in the NCBI's non-redundant database. The sequences that did not return an established vertebrate OR sequence as a 'best hit' were removed from further analyses. The sequences were shifted into the correct reading frame using a custom-written PERL script. Owing to the use of different primer pairs, OR fragments varied in length.

Thus, we restricted deduced receptor protein sequences to appropriate length (Freitag *et al.* 1998). Amplified avian partial OR-coding sequences were classified as being either non- γ -c or γ -c on the basis of sequence homologies between their corresponding predicted proteins and 78 potentially functional red jungle fowl (*G. gallus*) OR sequences of established classification (Niimura & Nei 2005). Note that the OR genes amplified with the primers annealing to the conserved regions ICKPLHY/NPFIYS(F/L) or VAICKPLHY/NPFIYS(F/L) that did not belong to the γ -c clade were removed from the analysis. A summary of all the amplified partial OR-coding sequences and the corresponding primer combinations used is shown in table S1 of the electronic supplementary material. We assigned a sequence as a potentially functional gene if an uninterrupted coding region was found (i.e. sequence without stop codon) while, if an interrupted coding region was found (i.e. sequence with stop codon), the sequence was assigned as a pseudogene (Gilad *et al.* 2004). In nine cases, copies of what appeared to be the same sequence were both potentially functional and pseudogenes and these were excluded from further analysis. Note that this method may overestimate the proportions of potentially functional OR genes, because frame-shift mutations outside the amplified coding region or mutations in promoter regions will not be detected. To determine how many potentially 'functional' OR-coding sequences from the experiments are indeed pseudogenes (owing to mutations outside the amplified region), we conducted a search for OR genes in the second draft of the *G. gallus* genome (BUILD v. 2.1, May 2006 release). The *G. gallus* OR sequences identified with the PCR-based method were then compared against the set of OR genes identified by the search using a BLAST approach. In addition, we compared the *G. gallus* sequences based on the degenerate PCR approach with Niimura & Nei's (2005) dataset, which was based on the first draft of the *G. gallus* genome.

A generalized linear mixed model (GLMM) was used to compare the proportion of potentially functional OR genes between the γ -c and the non- γ -c clades (Venables & Ripley 2002). The number of potentially functional OR genes amplified was used as the dependent variable, the total number of amplified OR genes as the binomial denominator, the species as a random factor and the clade as a predictor variable.

CLUSTAL X v. 1.81 (Thompson *et al.* 1997) was used with default parameters to construct multiple amino acid sequence alignments. The neighbour-joining (NJ) method was used to generate phylogenetic trees from Poisson correction distances using the MEGA software (<http://www.megasoftware.net/>). The reliability of the phylogenetic tree was evaluated with 1000 bootstrap repeats.

(c) Estimation of OR repertoire size

A non-parametric estimation technique applying the concept of 'sample coverage' (Chao & Lee 1992) was used to estimate the total number of OR genes in each of the nine avian genomes investigated. In a first step, the number of times identical PCR products were re-sequenced was used to estimate sample coverage (C) and its coefficient of variation (CV). In a second step, we chose the appropriate coverage estimator given the information provided by C and CV. This method does not assume an equal probability for each gene to be cloned and thus accounts for a primer bias. The black coucal was excluded from further analysis due to a large

CV. Abundance coverage estimators, their standard errors, confidence intervals and related statistics for all species were calculated using the software SPADE (<http://chao.stat.nthu.edu.tw/>) and can be found in table S2 of the electronic supplementary material. Note that the estimated total number of OR genes might be an underestimate of the true value (Bunge & Fitzpatrick 1993).

(d) Phylogenetically independent contrasts

To control for phylogenetic non-independence, we calculated phylogenetically independent contrasts (Felsenstein 1985) using the PDAP:PDTree module of MESQUITE (Midford *et al.* 2005; Maddison & Maddison 2006). The topology of the tree and branch lengths was obtained by using genetic distances derived from DNA-DNA hybridization studies (Sibley & Ahlquist 1991; see figure S1 of the electronic supplementary material). Since we could not estimate the number of OR genes from the black coucal (see §2c), we obtained seven contrasts from the eight species.

3. RESULTS

(a) Proportion of potentially functional OR genes

We amplified 46 distinct partial OR-coding sequences from red jungle fowl (*G. gallus*, order Galliformes) genomic DNA (table 1; table S1, electronic supplementary material). The large majority (95.7%) of the partial OR-coding sequences was predicted to be amplified from potentially functional OR genes. To determine whether this high potentially functional/non-functional ratio is a general characteristic of bird genomes, we amplified between 26 and 68 (mean \pm s.e.m. 53.5 ± 4.2) partial OR-coding sequences from a further eight species representing six additional avian orders (table 1; table S1, electronic supplementary material). The estimated proportion of potentially functional OR genes was consistently high in all taxa (mean \pm s.e.m. $83.7\% \pm 2.3\%$) despite the wide phylogenetic distribution and diverse ecological niches of the taxa examined (table 1). The estimated proportion of potentially functional OR genes was not statistically significantly different between the large γ -c OR clade (mean \pm s.e.m. $80.8\% \pm 3.9\%$) and the non- γ -c OR clade (mean \pm s.e.m. $85.7\% \pm 2.7\%$; GLMM, $t_{1,8} = 0.34$, $p = 0.74$).

(b) Comparison of data based on degenerate PCR and genome search

Of the 46 *G. gallus* sequences that we amplified using the degenerate PCR method, 18 were identical (more than or equal to 98.5% nucleotide identity) to the OR genes identified from the *G. gallus* genome search (BUILD v. 2.1, May 2006 release). The other 28 sequences were on average $94.9 \pm 0.5\%$ identical to the OR genes identified from the *G. gallus* genome search. Because the large majority of the other sequences (27 out of 28) mapped to 'chrUn_random' regions in the *G. gallus* genome and because the BUILD v. 2.1 genome draft still contains many sequence gaps, we assume that we amplified many OR-coding sequences that are not yet in the BUILD v. 2.1 genomic sequence.

A direct comparison of the results from the degenerate PCR and from the genome search showed that two coding sequences which were identified as potentially functional with the PCR-based method turned out to be pseudogenes due to mutations outside the amplified region. Thus, we overestimated the proportion of potentially

Table 1. Summary of data on relative olfactory bulb size (the OBR), the OR gene repertoire (the number of OR genes amplified, the number of pseudogenes amplified, the estimated percentage of potentially functional OR genes and the estimated total OR gene numbers) and ecological variables for the nine avian species.

common name	scientific name	order	OBR ^a	no. of OR genes amplified ^b	no. of pseudo-genes amplified ^b	percentage of potentially functional OR genes	estimated total number of OR genes	ecological variables ^c (habitat/diet/activity pattern)
canary	<i>Serinus canaria</i>	Passeriformes	6.0 ^d	55 (19/36/0)	11 (1/10/0)	80	166	scrub/granivorous/diurnal
galah	<i>Eolophus roseicapillus</i>	Psittaciformes	8.0 ^{d,e}	26 (17/9/0)	7 (3/4/0)	73.1	107	grassland and agricultural areas/herbivorous/diurnal
blue tit	<i>Cyanistes caeruleus</i>	Passeriformes	9.7 ^d	62 (18/43/1)	7 (6/1/0)	88.7	218	forest/omnivorous/diurnal
red jungle fowl	<i>Gallus gallus</i>	Galliformes	14.2 ^d	46 (11/35/0)	2 (0/2/0)	95.7	638	forest/omnivorous/diurnal
mallard	<i>Anas platyrhynchos</i>	Anseriformes	19.4 ^d	61 (14/47/1)	9 (0/9/0)	85.2	430	wetlands/omnivorous/diurnal
black coucal	<i>Centropus grillii</i>	Cuculiformes	19.5 ^{d,e}	61 (9/52/0)	8 (0/8/0)	86.9	—	grassland/insectivorous/diurnal
kakapo	<i>Strigops habroptilus</i>	Psittaciformes	30.2 ^f	56 (21/35/0)	10 (2/8/0)	82.1	667	forest/herbivorous/nocturnal
brown kiwi	<i>Apteryx australis</i>	Apterygiformes	34.0 ^d	68 (29/39/0)	16 (8/8/0)	76.5	600	forest and scrub/insectivorous/nocturnal
snow petrel	<i>Pagodroma nivea</i>	Procellariiformes	37.0 ^d	47 (21/26/0)	7 (3/4/0)	85.1	212	marine/planktivorous and piscivorous/diurnal

^aThe OBR is the ratio of the greatest diameter of the olfactory bulb relative to the greatest diameter of the cerebral hemisphere, expressed as a percentage.

^bNumbers in brackets refer to non- γ -c OR genes, γ -c OR genes and group- α OR genes, respectively.

^cFrom Cramp & Perrins (1992), del Hoyo *et al.* (1992, 1997, 2007), Elliot *et al.* (1994) and Bonadonna & Bretagnolle (2002).

^dFrom Bang & Cobb (1968).

^eMean OBR of this order.

^fFrom Hagelein (2004).

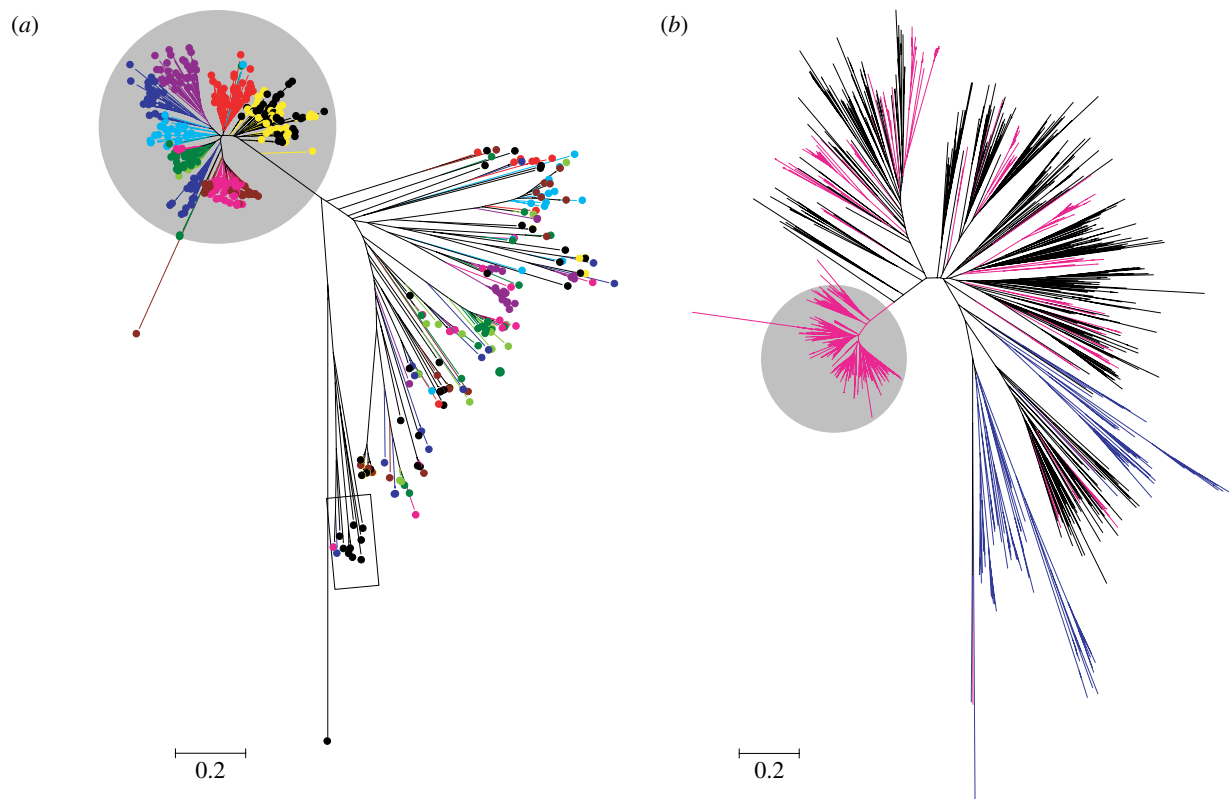


Figure 1. Unrooted neighbour-joining (NJ) phylogenetic trees derived from alignments of predicted vertebrate OR protein sequences (TM3–TM7). (a) The NJ phylogenetic tree of 483 predicted avian protein sequences derived from predicted functional OR genes from the canary (dark red, *S. canaria*, 44 sequences), the blue tit (pink, *C. caeruleus*, 55 sequences), the galah, (light green, *E. roseicapillus*, 19 sequences), the kakapo (dark green, *S. habroptilus*, 46 sequences), the black coucal (red, *C. grillii*, 53 sequences), the mallard (blue, *A. platyrhynchos*, 52 sequences), the red jungle fowl (yellow and black, *G. gallus*, 44 and 78 sequences), the snow petrel (cyan, *P. nivea*, 40 sequences) and the brown kiwi (purple, *A. australis*, 52 sequences). The red jungle fowl sequences that were obtained from Niimura & Nei (2005; $n=78$) are indicated by black circles, while the red jungle fowl sequences amplified in this study are indicated by yellow circles ($n=44$). Note that few group- α genes, indicated within the rectangle, were amplified using the primers and reaction conditions of this study. The large γ -c OR clade is shaded in grey. The scale bar indicates the number of amino acid substitutions per site. (b) Unrooted NJ trees generated from alignments of predicted vertebrate OR protein sequences: human (black lines, 388 sequences); zebrafish (blue lines, *Danio rerio*, 98 sequences); and avian (pink lines, 483 sequences). The predicted human and zebrafish OR protein sequences were obtained from Niimura & Nei (2005), while the avian OR sequences were from Niimura & Nei (2005; *G. gallus*, $n=78$) or this work. The γ -c OR clade is shaded in grey. The scale bar indicates the number of amino acid substitutions per site.

functional OR genes in the *G. gallus* genome by 11%. The comparison of the *G. gallus* sequences with Niimura & Nei's dataset yielded similar results (details not shown).

(c) Numbers of avian OR genes and relationship with relative olfactory bulb size

The estimated total number of OR genes, both potentially functional and non-functional, varied widely among the nine species investigated (range 107–667; table 1; table S2, electronic supplementary material). Our estimate for the *G. gallus*, 638 OR genes, is close to the previously reported estimate of 550 derived by Niimura & Nei (2005) from a draft *G. gallus* genome sequence. This suggests that our methodology provides a reasonably reliable estimate of OR gene repertoire sizes in those species for which full genomic sequences are not yet available. The lower values (107–218; canary, blue tit and galah) are within the range reported as typical of fishes (Ngai et al. 1993), while the higher values (600–667; red jungle fowl, brown kiwi and kakapo) rather resemble those of mammalian genomes (Glusman et al. 2001).

The estimated total number of OR genes, but not the proportion of potentially functional OR genes, correlated

positively with relative olfactory bulb size as measured by the OBR, the ratio of the greatest diameter of the olfactory bulb to the greatest diameter of the cerebral hemisphere in per cent (Bang & Cobb 1968; number: $r=0.63$, $n=8$, $p<0.05$ (one-tailed), figure S2a, electronic supplementary material; proportion: $r=0.20$, $n=9$, $p=0.6$, figure S2b, electronic supplementary material).

(d) Phylogenetic trees derived from predicted OR protein sequences

An expanded γ -c OR clade is present in all the nine avian genomes examined (figure 1a). This clade was supported with a high bootstrap value (91%). Within this clade, there is a strong tendency for sequences from the same species, or species from the same order, to cluster together (figure 1a). By contrast, among the non- γ -c OR sequences, the overall pattern is one of intermingling of sequences from differing taxa, presumably, reflecting that these gene lineages diverged before the diversification of these avian orders. An NJ tree based on an alignment of the 405 predicted potentially functional avian OR protein sequences identified in this study (table 1) and the corresponding regions of potentially functional OR proteins identified from the

red jungle fowl, zebrafish and human genome sequences (Niimura & Nei 2005) confirmed that the avian non- γ -c OR sequences intermingle with the other vertebrate OR protein sequences, whereas the avian γ -c OR clade sequences do not (figure 1b).

4. DISCUSSION

Our results strongly suggest that the proportion of potentially functional OR genes in avian genomes is considerably higher than the value of 15% estimated from an analysis of the BUILD v. 1.1 draft red jungle fowl (*G. gallus*) genomic sequence by Niimura & Nei (2005). Our results are consistent with those of the International Chicken Genome Sequencing Consortium (2004), who estimated that 93% (202 out of 218) of OR genes in the γ -c OR clade are potentially functional (also based on the analysis of the BUILD v. 1.1 draft red jungle fowl genome sequence). In addition, our results are consistent with an unpublished analysis of a more recent draft of the *G. gallus* genome (BUILD v. 2.1, May 2006 release; V. Kuryshv, S. S. Steiger, M. Stensmyr, B. Kempnaers & J. Mueller 2008, unpublished data). In this analysis, the proportion of potentially functional OR genes was estimated to be at least 53%. Note that even in the BUILD v. 2.1 draft, many putative partial OR genes still contain sequence gaps, so it is likely that many will be classified as functional in subsequent drafts. Assuming that all such incomplete OR genes are functional, the proportion of potentially functional OR genes in the *G. gallus* genome could be as high as 85% (V. Kuryshv, S. S. Steiger, M. Stensmyr, B. Kempnaers & J. Mueller 2008, unpublished data).

The estimated total number of OR genes differed widely between the bird genomes studied (range 107–667), indicating that different ecological niches may have shaped the OR gene repertoires in birds, as has been suggested for mammals (Niimura & Nei 2007). The observed differences in OR gene repertoire sizes are striking, but perhaps not too surprising for the following two reasons. First, birds also show a wide interspecific variation in the relative olfactory bulb size, as quantified by the OBR. For example, the OBR of the snow petrel (*P. nivea*) is 12 times larger than that of the black-capped chickadee (*Poecile atricapillus*; Bang & Cobb 1968). Hence, a similar interspecific variation in OR gene repertoire size could be expected. Second, in mammals, OR gene repertoire sizes range from 606 OR genes in the macaque to 2129 OR genes in the cow (Niimura & Nei 2007). Thus, OR gene repertoire sizes also greatly vary among mammalian species.

We estimated both the total number and the proportion of potentially functional OR genes in the nine different avian genomes using PCR primers annealing to evolutionarily conserved regions. Because it is unlikely that full genomic data for more avian species will become available in the near future (with the exception of the Australian passerine zebra finch, *Taeniopygia guttata*), PCR using degenerate primers is currently the only available method to study the avian OR gene repertoires in an ecological context. This method has already been used to estimate the fraction of potentially functional OR genes in relatively poorly characterized genomes of primates (Rouquier *et al.* 2000; Gilad *et al.* 2004),

carnivores (Quignon *et al.* 2003), rodents (Rouquier *et al.* 2000) and marine mammals (Kishida *et al.* 2007).

Notwithstanding its wide application, it is well recognized that this PCR-based approach has limitations and may overestimate the proportion of potentially functional OR genes, because (i) primer annealing sites may be more conserved in functional than in pseudogenes, and (ii) mutations that occur in regions not amplified by the primers will not be detected (Gilad *et al.* 2004). By comparing our PCR-based data with genome sequence information, we showed that the PCR-based approach overestimated the proportion of potentially functional OR-coding sequences in the *G. gallus* genome by approximately 11%. It is reasonable to assume that the extent of overestimation is similar for the other bird genomes.

Another disadvantage of the PCR-based method is that due to unpredictable primer bias, some OR genes may amplify preferentially. Thus, the ratios of partial OR-coding sequences among the amplification products may not represent a random sample of the OR repertoires in the genomes used as templates. However, if the primers were biased, we expect the bias to occur in all species and the between-species comparison should thus remain valid. Furthermore, it seems unlikely that a primer bias would generate a positive correlation between the estimated number of OR genes and the OBR. Finally, it has already been shown that PCR-based and whole-genome estimates yielded similar results. For example, Gilad *et al.* (2004) and Malnic *et al.* (2004) estimated the proportion of functional OR genes in humans to be approximately 48 and 53% in a PCR-based and genome-wide approach, respectively. Taken together, we argue that the PCR-based method is a useful approach to estimate the OR gene repertoires in birds.

Our results further suggest that estimating OR gene numbers in a wider range of avian genomes may provide insights into the selective pressures that have driven the evolution of avian olfaction. Ecological niche-associated adaptations such as daily activity pattern (e.g. nocturnal versus diurnal), habitat (e.g. terrestrial versus aquatic) or diet (e.g. generalist versus specialist) may well have shaped, and perhaps been driven by, the OR gene repertoires. For example, our finding that two night-active species, the kiwi and the kakapo, have comparatively large OR gene repertoires is consistent with the hypothesis that nocturnal species have evolved enhanced olfactory ability to deal with reduced effectiveness of vision under low-light conditions (Healy & Guilford 1990). The snow petrel seems to be an outlier in the sense that it has one of the largest OBRs measured in birds, but a relatively small estimated OR gene repertoire. However, in contrast to the kiwi and the kakapo, the snow petrel is a specialist diurnal forager (Ainley *et al.* 1984; Warham 1996) and it is plausible that its olfactory system has evolved to be highly sensitive to only a limited variety of odours. Based on our analysis, we predict that the OR gene repertoire of the zebra finch (*T. guttata*), whose genome sequence will soon become available, will be similar to that of the two passeriform genomes analysed here, approximately 200 OR genes.

We showed that OBR positively correlated with the estimated total number of OR genes, but not with the proportion of potentially functional OR genes, among the nine avian taxa examined. Thus, our results support the recent suggestion that the total number of OR genes, rather than the proportion of potentially functional

OR genes, is a correlate of olfactory ability (Niimura & Nei 2006). To account for phylogeny, we based our analysis on Sibley & Ahlquist's (1991) comprehensive, but somewhat controversial, topology. This phylogeny was used because it provides branch lengths, and including these greatly increased the power of the statistical analysis. However, note that when more recently suggested avian phylogenies lacking branch lengths were applied (Cracraft *et al.* 2004), the correlation of estimated OR gene number with OBR was no longer significant ($r=0.45$, $n=8$, $p=0.13$, one-tailed). Hence, an investigation of the OR gene repertoires of more avian species is needed to verify whether OBR is indeed positively correlated with OR gene repertoire size. It has been suggested that the size of the olfactory epithelium indicates olfactory ability (see Issel-Tarver & Rine (1997) and references therein). However, we could not test the correlation between the surface of the olfactory epithelium and OR gene repertoire size, because very little information exists about the surface of the olfactory epithelium in birds (Hagelin 2006). This may be worthy of future exploration.

While it is likely that birds with both relatively large OBRs and OR gene repertoires have an excellent sense of smell, the opposite may not be true. Thus, birds with relatively small OBRs and relatively few OR genes do not necessarily lack a good sense of smell. For example, despite their relatively small OBR (9.7%; Bang & Cobb 1968), European starlings (*Sturnus vulgaris*) are able to detect and discriminate volatile compounds of plants (e.g. milfoil *Achillea millefolium*) incorporated into their nests during the breeding season (Clark & Mason 1987; Gwinner & Berger 2008). Similarly, blue tits (*C. caeruleus*) appear to use olfaction in their maintenance of an aromatic environment for nestlings (Petit *et al.* 2002; Hagelin 2006) and for predator detection (Amo *et al.* 2008). Thus, the relationship between olfactory acuity, olfactory anatomy and OR gene repertoire characteristics is not simple and requires further study.

As a large γ -c OR clade is present in all the avian genomes examined, the γ -c OR clade expansion may be a characteristic of all the bird genomes. Two lines of evidence indicate that the γ -c OR clade expansion did not occur before the divergence of the avian lineage. First, we used the same degenerate PCR primer pairs to amplify OR-coding sequences from Nile crocodile (*Crocodylus niloticus*) genomic DNA and no γ -c OR genes were identified (S. Steiger 2007, unpublished data). Second, we did not detect any group- γ -c OR genes in database searches of a draft reptilian genomic sequence (*Anolis* lizard, *Anolis carolinensis*: V. Kuryshev 2008, unpublished data). Because the large γ -c OR clade is also absent from mammalian genomes, we suggest that this OR clade is a basal, shared feature of class Aves.

The red jungle fowl γ -c OR clade members were predicted to be orthologous to human OR genes located next to major histocompatibility complex (MHC) class I gene clusters (International Chicken Genome Sequencing Consortium 2004). Interestingly, MHC-linked OR genes may play a role in mating preferences (Penn 2002). Chicken MHC genes have been localized on microchromosome 16 (Fillon *et al.* 1996). However, to our knowledge, OR genes have not been located nearby. Since the majority of chicken OR genes have not been assigned positions on the genome (see below), it remains to be seen

whether avian γ -c OR clade members are in the proximity of MHC genes and/or are relevant for avian mate choice. Therefore, we suggest that future studies should investigate the functional significance of the apparently bird lineage-specific expanded γ -c OR clade.

The intermingling of the non- γ -c OR clade sequences of differing vertebrate taxa in the phylogenetic trees is compatible with the birth-and-death model of OR gene evolution, in which genes are created by repeated gene duplication and some genes later become non-functional (Nei 1969; for a review, see Nei & Rooney 2005). In addition, this pattern indicates that many of the OR gene lineage divergences pre-date the organism-level lineage divergences. Indeed, it is to be expected that a subset of the OR genes have evolutionarily conserved sequences and associated functions.

However, within the γ -c OR clade, sequences from the same, or closely related, species are very similar and therefore cluster together in phylogenetic trees. This clustering pattern may indicate that the γ -c OR clade may have arisen from independent expansion events or that the γ -c OR clade genes became homogenized by concerted evolution (Nei & Rooney 2005). Indeed, gene conversion has been shown to occur in closely related mammalian OR genes that are located together in a genomic cluster (Sharon *et al.* 1999). Interestingly, although the red jungle fowl γ -c OR genes have not yet been assigned to specific chromosomes (BUILD v. 2.1), BLAST searches have established that the 40 red jungle fowl γ -c OR genes identified by Niimura & Nei (2005) are located on 22 different contigs with a total length of 1691 kb. This represents only 0.14% of the total red jungle fowl genome, suggesting that the γ -c OR clade members may also be organized in clusters (data not shown). Such clustering promotes concerted evolution (Chen *et al.* 2007). Clearly, additional studies are needed to unravel both the molecular evolutionary history of the avian γ -c OR gene clade and its adaptive significance.

Available evidence suggests that OR genes with highly similar protein sequences bind structurally similar odorants (Malnic *et al.* 1999). If members of the large γ -c OR clade are functionally redundant, one would predict that loss-of-function mutations are not deleterious and, therefore, that a larger proportion of pseudogenes evolved in the γ -c OR clade than in the non- γ -c OR clades. However, the proportion of potentially functional OR genes does not differ significantly between the γ -c and non- γ -c OR clades, indicating that there is conservative or positive selection on genes forming the γ -c OR clade in all the avian genomes we examined.

In summary, our results support the growing body of evidence that the importance of the sense of smell for birds may have been greatly underestimated. In particular, the estimated OR gene repertoire sizes, and the proportion of OR genes that is potentially functional, contradict the general view that avian olfactory ability is poorly developed.

Blood collection procedures conformed to the animal experimental ethics regulations of the German Federal Republic, the European Union and New Zealand.

We thank Ursula Holter and Sylvia Kuhn for their expert laboratory assistance. We are grateful to Vladimir Kuryshev for his excellent biostatistical help. We thank Jakob Mueller,

James Dale, Nick Mundy, Alain Jacot and Kaspar Delhey for their helpful discussions, Naim Matasci and Henryk Milewski for writing and providing custom-written PERL codes and three anonymous reviewers for their constructive comments. Avian blood samples were generously provided by Francesco Bonadonna, CNRS-CEFE (snow petrel; financial and logistical support provided by the French Polar Institute IPEV), Angelika Denk, Max Planck Institute for Ornithology (MPIO, mallard), Wolfgang Goymann (MPIO, black coucal), Stefan Leitner (MPIO, canary), Hanne Lovlie, Stockholm University (red jungle fowl), Ron Moorhouse, New Zealand Department of Conservation (kakapo) and André Schuele, Berlin Zoo (brown kiwi). We thank Te Rūnanga o Ngāi Tahu (New Zealand) for their support of this research in allowing the use of kakapo DNA samples. We thank Matthias Stark for providing Nile crocodile tissue. International transport of DNA samples conformed to the legal requirements of the Convention on the International Trade of Endangered Species (CITES). This work was supported by the Max Planck Society.

REFERENCES

- Ainley, D. G., O'Connor, E. F. & Boekelheide, R. J. 1984 The marine ecology of birds in the Ross Sea, Antarctica. *Ornithol. Monogr.* **32**, 1–97.
- Amo, L., Galván, I., Tomás, G. & Sanz, J. J. 2008 Predator odour recognition and avoidance in a songbird. *Funct. Ecol.* **22**, 289–293. (doi:10.1111/j.1365-2435.2007.01361.x)
- Bang, B. G. & Cobb, S. 1968 Size of olfactory bulb in 108 species of birds. *Auk* **85**, 55–61.
- Bonadonna, F. & Bretagnolle, V. 2002 Smelling home: a good solution for burrow-finding in nocturnal petrels? *J. Exp. Biol.* **205**, 2519–2523.
- Bonadonna, F. & Nevitt, G. A. 2004 Partner-specific odor recognition in an Antarctic seabird. *Science* **306**, 835. (doi:10.1126/science.1103001)
- Buck, L. & Axel, R. 1991 A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187. (doi:10.1016/0092-8674(91)90418-X)
- Bunge, J. & Fitzpatrick, M. 1993 Estimating the number of species: a review. *J. Am. Stat. Assoc.* **88**, 364–373. (doi:10.2307/2290733)
- Chao, A. & Lee, S. M. 1992 Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87**, 210–217. (doi:10.2307/2290471)
- Chen, J.-M., Cooper, D. N., Chuzhanova, N., Ferec, C. & Patrinos, G. P. 2007 Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.* **8**, 762–775. (doi:10.1038/nrg2193)
- Clark, L. & Mason, J. R. 1987 Olfactory discrimination of plant volatiles by the European starling. *Anim. Behav.* **35**, 227–235. (doi:10.1016/S0003-3472(87)80228-2)
- Cracraft, J. *et al.* 2004 Phylogenetic relationships among modern birds (Neornithes). Toward an avian tree of life. In *Assembling the tree of life* (eds J. D. Cracraft & M. J. Donoghue), pp. 468–489. Oxford, UK: Oxford University Press.
- Cramp, S. & Perrins, C. M. (eds) 1992 *The birds of the Western Palearctic*, vol. VIII. *Crows to finches*. Oxford, UK: Oxford University Press.
- del Hoyo, J., Elliott, A. & Sargatal, J. 1992 *Handbook of the birds of the world: ostrich to ducks*. Barcelona, Spain: Lynx Edicions.
- del Hoyo, J., Elliott, A. & Sargatal, J. 1997 *Handbook of the birds of the world: sandgrouse to cuckoos*. Barcelona, Spain: Lynx Edicions.
- del Hoyo, J., Elliott, A. & Christie, D. A. 2007 *Handbook of the birds of the world: picathartes to tits and chickadees*. Barcelona, Spain: Lynx Edicions.
- Edinger, L. 1908 The relations of comparative anatomy to comparative psychology. *J. Comp. Neurol. Psychol.* **18**, 437–457. (doi:10.1002/cne.920180502)
- Elliott, A., Sargatal, J. & del Hoyo, J. 1994 *Handbook of the birds of the world: New World vultures to guinea fowl*. Barcelona, Spain: Lynx Edicions.
- Eriksson, J. *et al.* 2008 Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* **4**, e1000010. (doi:10.1371/journal.pgen.1000010)
- Felsenstein, J. 1985 Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15. (doi:10.1086/284325)
- Fillon, V., Zoorob, R., Yerle, M., Auffray, C. & Vignal, A. 1996 Mapping of the genetically independent chicken major histocompatibility complexes B° and $RFP-Y^{\circ}$ to the same microchromosome by two-color fluorescent *in situ* hybridization. *Cytogenet. Cell Genet.* **75**, 7–9. (doi:10.1159/000134445)
- Freitag, J., Ludwig, G., Andreini, I., Rössler, P. & Breer, H. 1998 Olfactory receptors in aquatic and terrestrial vertebrates. *J. Comp. Physiol. A* **183**, 635–650. (doi:10.1007/s003590050287)
- Freitag, J., Beck, A., Ludwig, G., von Buchholtz, L. & Breer, H. 1999 On the origin of the olfactory receptor family: receptor genes of the jawless fish (*Lampetra fluviatilis*). *Gene* **226**, 165–174. (doi:10.1016/S0378-1119(98)00575-7)
- Fuchs, T., Glusman, G., Horn-Saban, S., Lancet, D. & Pilpel, Y. 2001 The human olfactory subgenome: from sequence to structure and evolution. *Hum. Genet.* **108**, 1–13. (doi:10.1007/s004390000436)
- Gaillard, I., Rouquier, S. & Giorgi, D. 2003 Olfactory receptors. *Cell. Mol. Life Sci.* **61**, 456–469.
- Gilad, Y., Wiebel, V., Przeworski, M., Lancet, D. & Pääbo, S. 2004 Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* **2**, 120–125. (doi:10.1371/journal.pbio.0020005)
- Glusman, G., Yanai, I., Rubin, I. & Lancet, D. 2001 The complete human olfactory subgenome. *Genome Res.* **11**, 685–702. (doi:10.1101/gr.171001)
- Gwinner, H. & Berger, S. 2008 Starling males select green nest material by olfaction using experience-independent and experience-dependent cues. *Anim. Behav.* **75**, 971–976.
- Hagelin, J. C. 2004 Observations on the olfactory ability of the kakapo *Strigops habroptilus*, the critically endangered parrot of New Zealand. *Ibis* **146**, 161–164. (doi:10.1111/j.1474-919X.2004.00212.x)
- Hagelin, J. C. 2006 Odors and chemical signaling. In *Reproductive behavior and phylogeny of birds: sexual selection, behavior, conservation, embryology and genetics* (ed. B. G. M. Jamieson), pp. 75–119. Enfield, NH: Science Publishers.
- Hagelin, J. C. & Jones, I. L. 2007 Bird odors and other chemical substances: a defense mechanism or overlooked mode of intraspecific communication? *Auk* **124**, 1–21. (doi:10.1642/0004-8038(2007)124[74:BOA OCS]2.0.CO;2)
- Hall, T. A. 1999 BioEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
- Healy, S. & Guilford, T. 1990 Olfactory-bulb size and nocturnality in birds. *Evolution* **44**, 339–346. (doi:10.2307/2409412)
- International Chicken Genome Sequencing Consortium 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716. (doi:10.1038/nature03154)
- Issel-Tarver, L. & Rine, J. 1997 The evolution of mammalian olfactory receptor genes. *Genetics* **145**, 185–195.
- Kishida, T., Kubota, S., Shirayama, Y. & 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**, 428–430. (doi:10.1098/rsbl.2007.0191)
- Lagerström, M. C., Hellström, A. R., Gloriam, D. E., Larsson, T. P., Schiöth, H. B. & Fredriksson, R. 2006 The G protein—coupled receptor subset of the chicken genome. *PLoS Comput. Biol.* **2**, 493–507. (doi:10.1371/journal.pcbi.0020054)
- Leibovici, M., Lapointe, F., Aletta, P. & Ayer-Le Lièvre, C. 1996 Avian olfactory receptors: differentiation of olfactory neurons under normal and experimental conditions. *Dev. Biol.* **175**, 118–131. (doi:10.1006/dbio.1996.0100)
- Maddison, W. P. & Maddison, D. R. 2006 *MESQUITE: a modular system for evolutionary analysis*, v. 1.12. See <http://mesquiteproject.org>.
- Malnic, B., Hirono, J., Sato, T. & Buck, L. 1999 Combinatorial receptor codes for odors. *Cell* **96**, 713–723. (doi:10.1016/S0092-8674(00)80581-4)
- Malnic, B., Godfrey, P. A. & Buck, L. B. 2004 The human olfactory receptor gene family. *Proc. Natl Acad. Sci. USA* **101**, 2584–2589. [Erratum in *Proc. Natl Acad. Sci. USA* 2004 **101**, 7205.] (doi:10.1073/pnas.0307882100)
- Midford, P. E., Garland Jr, T. & Maddison, W. P. 2005 PDAP package of MESQUITE, v. 1.07.
- Mombaerts, P. 2004 Genes and ligands for odorant, vomeronasal and taste receptors. *Nat. Rev. Neurosci.* **5**, 263–278. (doi:10.1038/nrn1365)
- Nef, S., Allaman, I., Fiumelli, H., De Castro, E. & Nef, P. 1996 Olfaction in birds: differential embryonic expression of nine putative odorant receptor genes in the avian olfactory system. *Mech. Dev.* **55**, 65–77. (doi:10.1016/0925-4773(95)00491-2)
- Nei, M. 1969 Gene duplication and nucleotide substitution in evolution. *Nature* **221**, 40–42. (doi:10.1038/221040a0)
- Nei, M. & Rooney, A. P. 2005 Concerted and birth-and-death evolution of multigene families. *Ann. Rev. Genet.* **39**, 121–152. (doi:10.1146/annurev.genet.39.073003.112240)
- Nei, M., Gu, X. & Sitnikova, T. 1997 Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl Acad. Sci. USA* **94**, 7799–7806. (doi:10.1073/pnas.94.15.7799)
- Nevitt, G. A., Losekoot, M. & Weimerskirch, H. 2008 Evidence for olfactory search in wandering albatross, *Diomedea exulans*. *Proc. Natl Acad. Sci. USA* **105**, 4576–4581. (doi:10.1073/pnas.0709047105)
- Ngai, J., Dowling, M. M., Buck, L., Axel, R. & Chess, A. 1993 The family of genes encoding odorant receptors in the channel catfish. *Cell* **72**, 657–666. (doi:10.1016/0092-8674(93)90396-8)
- Niimura, Y. & Nei, M. 2005 Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc. Natl Acad. Sci. USA* **102**, 6039–6044. (doi:10.1073/pnas.0501922102)
- Niimura, Y. & Nei, M. 2006 Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J. Hum. Genet.* **51**, 505–517. (doi:10.1007/s10038-006-0391-8)
- Niimura, Y. & Nei, M. 2007 Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* **2**, e708. (doi:10.1371/journal.pone.0000708)
- Papi, F. 1991 Olfactory navigation. In *Orientation in birds* (ed. P. Berthold), pp. 52–85. Basel, Switzerland: Birkhauser Verlag.
- Penn, D. J. 2002 The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology* **108**, 1–21. (doi:10.1046/j.1439-0310.2002.00768.x)
- Petit, C., Hossaert-McKey, M., Perret, P., Blondel, J. & Lambrechts, M. M. 2002 Blue tits use selected plants and olfaction to maintain an aromatic environment for nestlings. *Ecol. Lett.* **5**, 585–589. (doi:10.1046/j.1461-0248.2002.00361.x)
- Quignon, P. et al. 2003 Comparison of the canine and human olfactory receptor gene repertoires. *Genome Biol.* **4**, R80. (doi:10.1186/gb-2003-4-12-r80)
- Roper, T. J. 1999 Olfaction in birds. *Adv. Study Behav.* **28**, 247–332.
- Rouquier, S., Blancher, A. & 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. USA* **97**, 2870–2874. (doi:10.1073/pnas.040580197)
- Sharon, D., Glusman, G., Pilpel, Y., Khen, M., Gruetzner, F., Haaf, T. & Lancet, D. 1999 Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics* **61**, 24–26. (doi:10.1006/geno.1999.5900)
- Sibley, C. G. & Ahlquist, J. E. 1991 *Phylogeny and classification of birds: a study in molecular evolution*. New Haven, CT: Yale University Press.
- Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F. & Higgins, D. 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882. (doi:10.1093/nar/25.24.4876)
- Venables, W. N. & Ripley, D. 2002 *Modern applied statistics with S*. Berlin, Germany: Springer.
- Warham, J. 1996 *The behaviour, population biology and physiology of the petrels*. London, UK: Academic Press.
- Wenzel, B. M. 1968 Olfactory prowess of kiwi. *Nature* **220**, 1133–1134. (doi:10.1038/2201133a0)
- Young, J. M. & Trask, B. J. 2002 The sense of smell: genomics of vertebrate odorant receptors. *Hum. Mol. Genet.* **11**, 1153–1160. (doi:10.1093/hmg/11.10.1153)