Evolution of avian olfaction

DISSERTATION

der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

vorgelegt von

Silke S. Steiger

München, Mai 2008

durchgeführt am Max Planck Institut für Ornithologie
in Seewiesen

CONTENTS

Chapter **1**

General introduction

The sense of smell is of central importance to most animals because it can be used, for example, to locate food, to navigate or to avoid predators. Olfactory cues can also play an important role for communication. The molecular basis of the sense of smell is mediated by olfactory receptors (ORs) expressed on sensory neurons in the olfactory epithelium (Buck and Axel, 1991). The activation of ORs by volatile odorants (i.e., small organic molecules such as various alcohols, aliphatic acids, aldehydes, ketones, and esters) represents the first step of a transduction cascade that finally enables odour detection (for review, see Firestein, 2001). Notably, Richard Axel and Linda Buck were awarded the Nobel Prize in Physiology/Medicine for their discoveries of odorant receptors and the organization of the olfactory system in 2004.

Since their discovery, OR genes have been intensively studied in a wide range of vertebrates, from fish to mammals (for reviews, see Mombaerts, 1999a; Zhang and Firestein, 2002; Niimura and Nei, 2006). However, in contrast, OR genes in birds have been mostly ignored (Figure 1.1). This is most likely because it is still widely believed that birds lack a well-developed sense of smell, despite the fact that a functional olfactory system has been demonstrated in every bird species studied so far (for reviews, see Roper, 1999; Hagelin, 2006; Hagelin and Jones, 2007). This thesis aims to provide more insights into the genetic basis of the sense of smell in birds and the evolution of avian chemoreception.
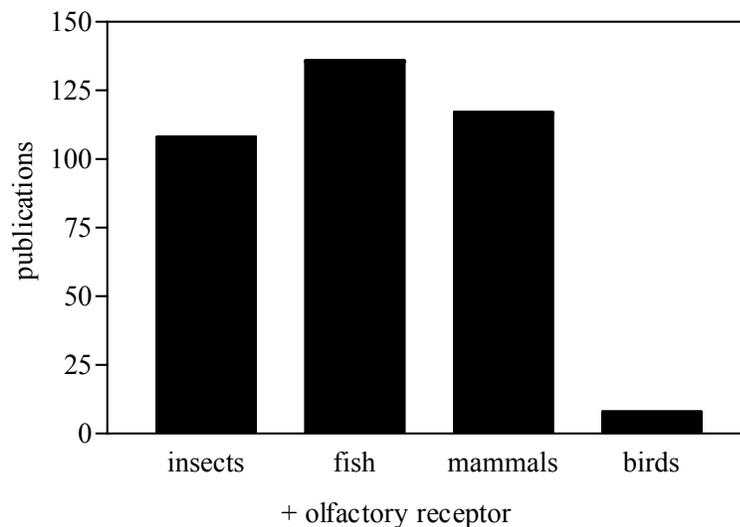
Figure 1.1
Results of a Web of Knowledge based literature search (ISI Web of Knowledge, http://isiwebofknowledge.com). Respective taxons (insects, fish, mammals, birds) in combination with the term 'olfactory receptor' were used as search term. The years from 1900 - 2008 were selected as timespan.

In the remainder of this Chapter, I summarize the current knowledge about vertebrate chemosensory receptors and the avian sense of smell. In addition, I briefly describe the contents of the remaining Chapters of this thesis (Chapter 2-7).

OLFACTORY RECEPTOR GENES

OR genes are small (~1 kb) and intronless (Buck and Axel, 1991). They represent a large gene family in vertebrate genomes (for reviews, see Gaillard et al., 2004; Mombaerts, 2004). The size of the OR gene family varies widely between vertebrate genomes (size range: 100 - 2130 in the pufferfish, *Fugu rubripes*, and the cow, *Bos taurus*, respectively) (Niimura and Nei, 2006; Niimura and Nei, 2007). Further, it is known that a fraction of the OR gene repertoire has degenerated to pseudogenes. Pseudogenes are sequences that are similar to one or more paralogous genes but that have lost their protein-coding ability due to mutations and thus, are non-functional (Mighell et al., 2000). Interestingly, the proportion of pseudogenes also varies widely between genomes. In mammals, the predicted proportion of OR genes that are pseudogenes ranges from 12% in dog, *Canis lupus familiaris*, to 50-60% in humans, *Homo sapiens* (Mombaerts, 2004; Niimura and Nei, 2006; Niimura and Nei, 2007) (Figure 1.2).
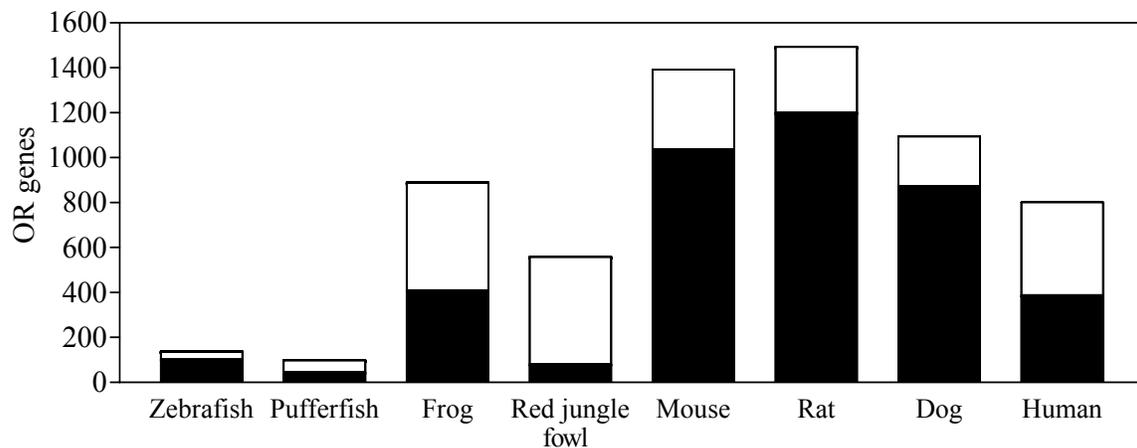
Figure 1.2
Numbers of functional olfactory receptor (OR) genes (black bars) and pseudogenes (white bars) for different vertebrate species (zebrafish, *Danio rerio*; pufferfish, *Fugu rubripes*; frog, *Xenopus tropicalis*; red jungle fowl, *Gallus gallus*; mouse, *Mus musculus*; rat, *Rattus norvegicus*; dog, *Canis lupus familiaris*; human, *Homo sapiens*; adapted from Niimura and Nei, 2006). Note that the proportion of functional OR genes for the red jungle fowl is most likely highly underestimated (Niimura and Nei, 2005).

Comparative genomic studies suggested 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; but see Laska et al., 2005). The total number of OR genes in a genome may reflect how many different scents can be detected and distinguished (Niimura and 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 and 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 to other primates (Rouquier et al., 2000; Gilad et al., 2004).

It should be noted that olfactory chemoreceptors have also been identified in invertebrates (e.g. in the nematode *Caenorhabditis elegans* and the fruit fly, *Drosophila*). However, sequence identities of vertebrate and non-vertebrate olfactory chemoreceptors are very low (Dahanukar et al., 2005). For example, *Caenorhabditis elegans* chemoreceptors share only ~10% sequence identity with vertebrate OR genes. In addition, non-vertebrate olfactory chemoreceptors contain introns. As a result, it has been questioned whether non-vertebrate and vertebrate OR genes derive from a common ancestor (Gaillard et al., 2004).

In situ hybridization (FISH) and database mining approaches demonstrated that OR genes occur in clusters in vertebrate genomes. For example, in the human genome, 95 OR genomic clusters (6-138 genes each) could be identified on all chromosomes except chromosomes 20 and Y (Glusman et al., 2001; Niimura and Nei, 2003). Similarly, OR genes were assigned to all mouse chromosomes except chromosomes 5, 12, 18, and the Y chromosome (Godfrey et al., 2004). OR genes are not equally distributed on the chromosomes. For example, six chromosomes account for ~70% of the human OR gene repertoire (Glusman et al., 2001). Based on these findings, it has been suggested that tandem gene duplications are involved in OR family expansion (Young et al., 2002).

Vertebrate OR genes have been classified into two distinct groups based on sequence identities: class-α (also termed class I genes) and class-γ (also termed class II genes) (Freitag et al., 1995; Niimura and Nei, 2005). Class-α genes have been first identified in fish and thus, were thought to activated by water-soluble odorants in an aquatic environment. As class-γ genes were predominantly found in mammals, they were believed to be activated by volatile odorants in a terrestrial environment (Freitag et al., 1995; Freitag et al., 1998). OR genes can be further classified into families and subfamilies, based on amino acid sequence identity (family: >40% identity, subfamily: >60% identity). For example, human class-α genes were divided into 17 families and class-γ genes into 14 families (Glusman et al., 2001). In total, the human OR family is composed of 172 subfamilies (Malnic et al., 2004). Members of the same subfamily are very similar in sequence and thus, are likely to recognize structurally related odorants (Malnic et al., 2004). Interestingly, many OR genes that belong to one phylogenetic clade were located in the same genomic cluster, supporting the idea that tandem gene duplication has been an important mechanism in the evolution of OR genes (Niimura and Nei, 2003).

Little is currently known about avian OR genes (Figure 1.1), probably because it is still widely believed that birds lack a well-developed sense of smell (for review, see Roper, 1999). The total number of OR genes in the red jungle fowl (*Gallus gallus*) was recently estimated to be between 283 (International Chicken Genome Sequencing Consortium, 2004) and 558 (Niimura and Nei, 2005), and the first estimates of the number of functional OR genes ranged from 78 (Niimura and Nei, 2005) to 229 (Lagerstrom et al., 2006). Both studies were based on a data mining approach of the red jungle fowl genome

draft (Build 1.1, released in February 2004; Figure 1.3). The substantial lack of agreement in the estimated number of OR genes (both in total and functional) is surprising but can most likely be attributed to the different bioinformatics search strategies used. Interestingly, the 'large size of the olfactory receptor family seems to run counter to the textbook view that birds have a poor sense of smell' (International Chicken Genome Sequencing Consortium, 2004). Notably, a subset of the class-γ OR genes, termed class-γ-c, is greatly expanded in the red jungle fowl genome (Niimura and Nei, 2005; Lagerstrom et al., 2006). So far, nothing is known about OR genes in other bird species than the red jungle fowl.

OLFACTORY RECEPTORS

Vertebrate OR genes are expressed in the olfactory epithelium on the surface of the cilia of the olfactory sensory neurons. Each olfactory sensory neuron only expresses one OR gene (Mombaerts, 2004). Axons from all sensory neurons that express that particular receptor merge into so called 'glomeruli' in the olfactory bulb. From the olfactory bulb, signals are delivered and finally processed to higher brain structures such as the piriform cortex, the hippocampus and the amygdala (Mombaerts, 2004), where they result in the perception of smells.

Olfactory receptors are members of a large family of seven-transmembrane (TM)-domain G-protein coupled receptors (GPCRs) and are about 300-350 amino acids long (for review, see Gaillard et al., 2004). The TM regions are connected by three intracellular (IC) and three extracellular (EC) loops. Notably, EC 2 is particularly long in ORs (Figure 1.4). Olfactory receptors possess both highly conserved and hypervariable protein regions. Strong sequence diversity can be particularly found in the third, fourth and fifth TM region. These TM regions are most likely involved in odorant binding (Floriano et al., 2004). Highly conserved amino acid motifs within and across species such as MAYDRYVAIC at the end of TM3 can be used to classify a sequence as an OR protein (Young and Trask, 2002; Liu et al., 2003).

Vertebrates are able to perceive and discriminate many more different volatile chemicals than there are ORs encoded in the genome. The detection and discrimination of such a vast number of volatile chemicals is most likely achieved through a mechanism known as the 'combinatorial receptor code' (Malnic et al., 1999). Briefly, one odour molecule can be recognized by several olfactory receptors, and one olfactory receptor can recognize several odour molecules.

Interestingly, OR genes are not only expressed in the olfactory epithelium. Instead, transcripts of a subset of OR genes have been found in testis and sperm and other non-olfaction related tissues of both mammals and fish (Parmentier et al., 1992; Vanderhaeghen et al., 1993; Vanderhaeghen et al., 1997; Spehr et al., 2003), suggesting that the function of ORs is not restricted to the context in which they were first characterized, namely olfaction (for review, see Young and Trask, 2002). Interestingly, mammalian testicular/sperm OR proteins have been localized in the sperm flagellum midpiece and have been shown to mediate flagellar motility (Spehr et al., 2003). Hence, it is plausible that sperm expressed OR genes play a role in sperm-egg chemotaxis/communication (Spehr et al., 2006).

Figure 1.3 (opposite page)
Phylogeny of red jungle fowl (*Gallus gallus*) olfactory receptors (OR). A Neighbour-Joining (NJ) tree was constructed based on an alignment of predicted amino acid sequences of 78 putatively functional ORs identified from the red jungle fowl genome draft (Build 1.1). Sequences were obtained from Niimura and Nei (2005). As outgroup, the chicken melanocortin receptor1 (GenBank accession number **S71420**) was used. Numbers at internal nodes indicate branches supported with >90% bootstrap values. The scale bar indicates the estimated number of amino acid substitutions per site. Clades defined by Niimura and Nei (2005) are shown. Alternative nomenclature is shown in brackets. The program MEGA (Kumar et al., 2004; http://www.megasoftware.net/) was used with default settings to construct the phylogenetic tree.

MOLECULAR EVOLUTION OF OLFACTORY RECEPTOR GENES

What selective pressures have shaped the OR gene repertoires? Selective pressures acting at the protein level are typically assessed by determining the ratio of nonsynonymous versus synonymous mutations ($\omega = d_N/d_S$) (Yang and Bielawski, 2000). A nonsynonymous mutation leads to a change in the amino acid, whereas a synonymous mutation does not. Positive selection, that is selection leading to changes in the protein sequence, is indicated by $\omega > 1$. Purifying selection, that is selection favoring stability in the protein sequence, is indicated by $\omega < 1$. Neutral selection is indicated by $\omega = 1$. The ω ratio has either been calculated as an average over all codon sites (Nei and Gojobori, 1986; Yang and Nielsen, 2000) or on a site-by site basis (i.e. at every codon) (Nielsen and Yang, 1998; Suzuki and Gojobori, 1999) in the sequence region of interest.

The striking diversity of OR genes has been associated with codon site-specific positive selection (i.e. $\omega > 1$). For example, previous studies found evidence for positive selection in fish (e.g. Ngai et al., 1993; Alioto and Ngai, 2005) and mammalian ORs (e.g. Hughes and Hughes, 1993; Gilad et al., 2003). As expected, positively selected codons were predominantly located in TM regions that are most likely involved in ligand binding. However, other studies did not find any evidence for positively selected codon sites (Gimelbrant et al., 2004; Zhang et al., 2004b). Thus, further investigation is needed to determine whether positive selection has played an important role in the evolution of vertebrate ORs (Niimura and Nei, 2006). To date, selective pressures acting on avian ORs have not been studied.

METHODS USED TO INVESTIGATE OLFACTORY RECEPTOR GENES

Generally, three methods have been used to identify OR genes: (i) polymerase chain reaction (PCR) amplification using degenerate oligonucleotide primers that anneal to evolutionary conserved protein regions, (ii) Southern Blot hybridisation to estimate the size of OR subfamilies and (iii) genome database searches. For the purpose of this thesis, all three methods were used, and I discuss them here briefly.

TM1	TM2	TM3	TM4	TM5	TM6	TM7

	IC1	EC1	IC2	EC2	IC3	EC3
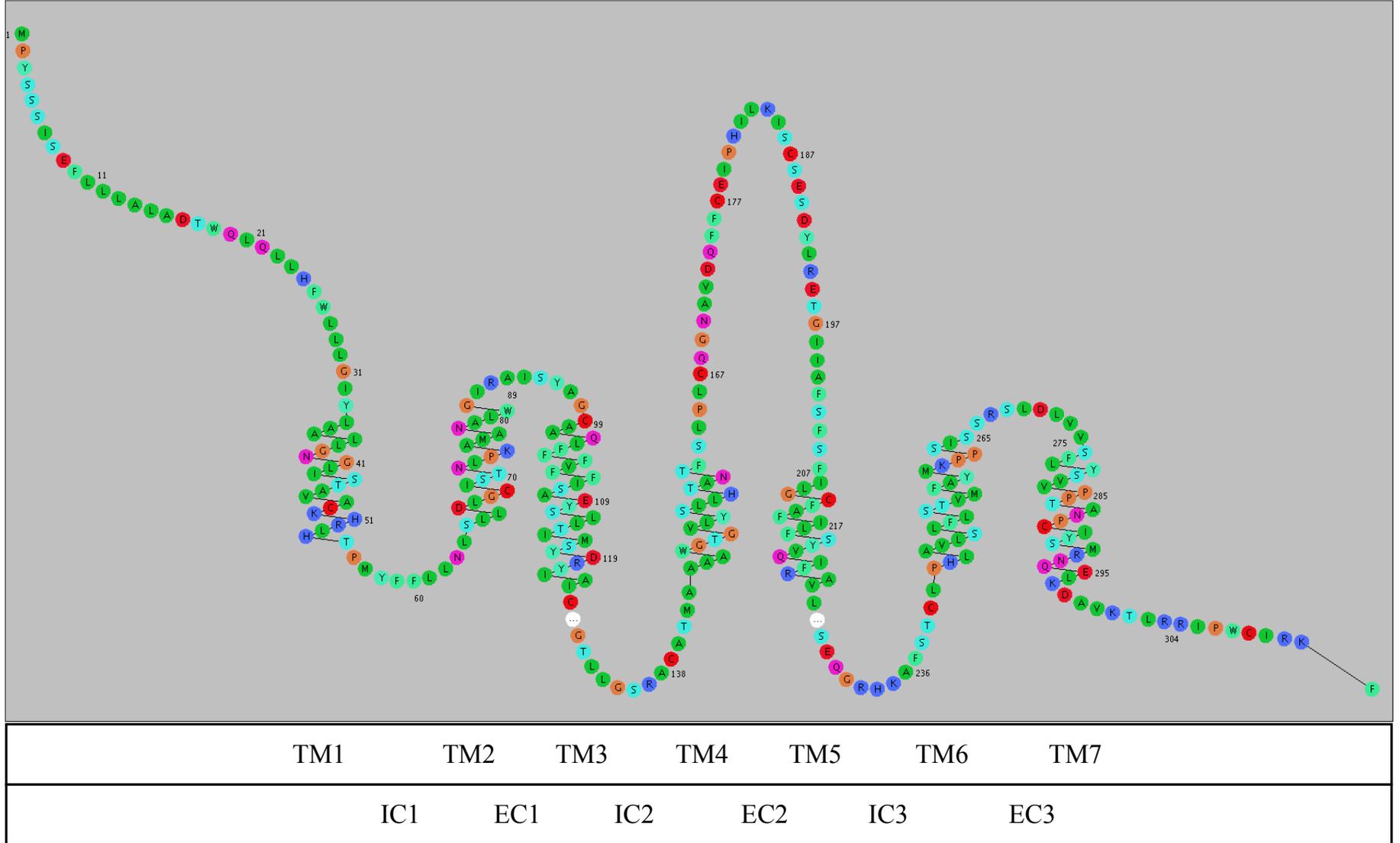
Figure 1.4 (opposite page)

A schematic representation of one red jungle fowl (*Gallus gallus*) olfactory receptor (OR) identified by Niimura and Nei (2005; Gg2ORUn.21). The RbDe web service was used to generate the snake plot (http://icb.med.cornell.edu/crt/RbDe/RbDe.html). Each circle represents one amino acid. ORs are seven-transmembrane (TM) proteins. TM domains were predicted using the BCM Search Launcher (http://searchlauncher.bcm.tmc.edu/seq-search/struc-predict.html). Note that the exact number and precise placement of the TM domains has not been experimentally verified and thus, is speculative. Further note that some residues are hidden from the diagram because of lack of space (indicated by white circles with three dots). Boxes below the snake plot indicate the position of TM domains and intracellular (IC) and extracellular (EC) loops, respectively.

(a) PCR amplification

PCR using degenerate primers (degenerate primers are a mixture of similar, but not identical primers) has proven to be a useful method to identify novel members of OR gene families and to determine the functional/non-functional OR gene ratio. Degenerate primers were designed to anneal to highly conserved OR regions, that usually are both conserved within and across species (e.g. to evolutionary conserved TM1, TM2, TM3, and TM7 regions (Buck and Axel, 1991). For example, Gilad et al. (2004) used primer pairs designed to amplify a 670 bp product that approximately covers the region from TM2–7 of the OR protein. Using degenerate primers, several OR genes can be amplified in a single PCR reaction from genomic DNA. Subsequent cloning and sequencing of the PCR products allows an estimation of the proportion of functional and nonfunctional OR genes within a given genome. This method has been successfully used to determine the functional/nonfunctional OR gene ratio in mammals (e.g. primates (Rouquier et al., 2000; Gilad et al., 2004), carnivores (Quignon et al., 2003), rodents (Rouquier et al., 2000) and cetaceans (Kishida et al., 2007)).

It should be noted that the PCR method has several drawbacks. First, due to unpredictable primer bias, some OR genes may amplify preferentially. Thus, the OR partial coding sequences amongst the amplification products may not represent a random sample of the OR genes in the investigated genomes. Second, degenerate primers usually do not amplify the complete reading frame of an OR gene. Mutations that occur in regions not amplified by the primers (including promoter or control regions of OR genes) will not be detected and therefore the proportion of functional OR genes may be overestimated. Nevertheless, degenerate PCR is a useful method to study OR genes in animals in species for which full genomic sequences are not yet available.

(b) Southern blot hybridization

Southern blot hybridizations using unique sequence probes that represent specific OR subfamilies have been used to estimate the sizes of OR gene subfamilies in fish (Sun et al., 1999; Irie-Kushiyama et al., 2004), amphibians (Freitag et al., 1998), birds (i.e. chicken, Leibovici et al., 1996; Nef and Nef, 1997) and mammals (Issel-Tarver and Rine, 1996; Issel-Tarver and Rine, 1997). Although Southern hybridization has the advantage of directly investigating the entire genome, it does not provide information on the functional status of the hybridizing sequences.

(c) Genome searches

Entire OR gene repertoires have been identified in species for which full genomic sequences are available (e.g. red jungle fowl, mouse, dog, human) by database mining approaches using bioinformatics techniques (for review, see Niimura and Nei, 2006). For example, a TBLASTN search strategy was used to identify mouse OR genes (Godfrey et al., 2004). It is important to note that different gene search algorithms may yield very different results: for example, Niimura and Nei (2005) identified 78 functional OR genes whereas Lagerstrom et al. (2006) identified 229 functional OR genes in the red jungle fowl, even though both studies used the same version of the red jungle fowl genome (Build 1.1).

ANOTHER CLASS OF CHEMOSENSORY RECEPTORS: TAAR GENES

Recently, a second class of chemosensory receptors, trace amine-associated receptors (TAARs), have been described in the mouse olfactory epithelium (Gloriam et al., 2005; Liberles and Buck, 2006). Like ORs, TAARs are members of G-protein coupled receptors. Interestingly, the TAAR gene repertoire also varies tremendously among vertebrates (range: 5-37 in human and zebrafish, respectively) (Gloriam et al., 2005; Hashiguchi and Nishida, 2007). Natural ligands identified for mouse TAARs have been detected in mouse urine which is known to be a major source of social cues (Liberles and Buck, 2006). Therefore, it has been suggested that TAARs may be highly relevant for social communication and individual recognition.

THE AVIAN SENSE OF SMELL

Historically, birds were thought to rely heavily on visual and auditory signals, whereas olfactory signals were predicted to be of minor importance (for review, see Roper, 1999). Therefore, it is not surprising that the study of the sense of smell in birds has received scant attention. Nevertheless, highly valuable information has been collected about variance in the size of the olfactory bulb, a morphological correlate of olfactory ability (Edinger, 1908). The olfactory bulb size ratio (OBR), measured as the ratio of the greatest diameter of the olfactory bulb to the greatest diameter of the cerebral hemisphere (expressed in per cent), was investigated in more than 150 species from 23 orders (Bang and Cobb, 1968; Bang, 1971; Bang and Wenzel, 1985), including highly endangered species such as the nocturnal kakapo (*Strigops habroptilus*) (Hagelin, 2004). Interestingly, there is striking variation among avian relative olfactory bulb sizes. OBR ranges from only 3% in the black-capped chickadee (*Parus atricapillus*) up to 37% in the snow petrel (*Pagodroma nivea*) (Bang and Cobb, 1968). A comparative study showed that nocturnal and crepuscular species have larger OB sizes than their diurnal relatives, suggesting a relationship between ecology and the evolutionary development of a part of the brain (Healy and Guilford, 1990).

In recent years, it has been shown that birds can use the sense of smell in a variety of contexts such as navigation (Papi, 1991; Wallraff, 2004), foraging (Nevitt et al., 2008), predator detection and avoidance (Amo et al., 2008), nest-building (Petit et al., 2002; Gwinner and Berger, 2008) and partner recognition (Bonadonna and Nevitt, 2004). Although birds in the order Passeriformes were long thought to lack a well developed sense of smell due to the small size of their olfactory bulb, there is now evidence that some passerines are quite well able to detect odours (e.g. Clark and Smeraski, 1990;

Petit et al., 2002; Gwinner and Berger, 2008) and to use olfaction for predator detection (Amo et al., 2008). In addition, songbirds most likely use olfactory cues for migration (Martin Wikelski, personal communication). Taken together, recent evidence from behavioural and neurophysiological studies strongly suggests that the sense of smell in birds may be a more important sense than previously thought. Although the avian sense of smell has been investigated in several behavioural, anatomical and neurophysiological studies (for review, see Roper, 1999; Hagelin, 2006; Hagelin and Jones, 2007), the molecular basis of olfaction, namely OR genes, has hitherto hardly been studied (Figure 1.1).

OUTLINE OF THIS THESIS

In this thesis, I investigate OR gene repertoires in a variety of bird species (Table 1.1). First, I determine the proportion of functional OR genes and estimate the OR gene repertoire sizes in nine avian species from seven different orders using degenerate PCR with primers specific for the OR gene family (Chapter 2).

Table 1.1
Bird species studied in this thesis.

Common name	Latin name	Order	Family	Chapter
Black coucal	*Centropus grillii*	Cuculiformes	Cuculidae	2
Blue tit	*Cyanistes caeruleus*	Passeriformes	Paridae	2
Brown kiwi	*Apteryx australis*	Apterygiformes	Apterygidae	2,3
Canary	*Serinus canaria*	Passeriformes	Fringillidae	2
(Domestic) Chicken	*Gallus gallus domesticus*	Galliformes	Phasianidae	5
Elegant-crested tinamou	*Eudromia elegans*	Tinamiformes	Tinamidae	3
Emu	*Dromaius novaehollandiae*	Casuariiformes	Dromaiidae	3
Galah	*Eolophus roseicapillus*	Psittaciformes	Cacatuidae	2,3
Greater Rhea	*Rhea americana*	Rheiformes	Rheidae	3
Red jungle fowl	*Gallus gallus*	Galliformes	Phasianidae	2,3,4,5,6
Kaka	*Nestor meridionalis*	Psittaciformes	Psittacidae	3
Kakapo	*Strigops habroptilus*	Psittaciformes	Psittacidae	2,3
Kea	*Nestor notabilis*	Psittaciformes	Psittacidae	3
Mallard	*Anas platyrhynchos*	Anseriformes	Anatidae	2
Ostrich	*Struthio camelus*	Struthioniformes	Struthionidae	3
Snow petrel	*Pagodroma nivea*	Procellariiformes	Procellariidae	2

Second, I use a Southern Blot approach in combination with a PCR based approach to compare the OR gene repertoires of closely related bird species. I ask whether ecological variables such as activity patterns shaped the evolution of OR genes (Chapter 3). Third, I explore whether avian OR genes show signatures of positive selection (Chapter 4). Fourth, I examine whether OR genes are transcribed in domestic chicken (*Gallus gallus domesticus*) testes (Chapter 5) using Reverse Transcription (RT)-PCR. Fifth, I investigate whether TAAR orthologues are encoded in the red jungle fowl genome using a BLAST search strategy (Chapter 6). Finally, I summarize and discuss the results and suggest some directions for future research (Chapter 7).

Chapter 2

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

Silke S. Steiger, Andrew E. Fidler, Mihai Valcu and Bart Kempenaers

ABSTRACT

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 petrel, *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 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.

INTRODUCTION

Olfactory receptors (OR) expressed in sensory neurons within the olfactory epithelium constitute the molecular basis of the sense of smell among vertebrates (Buck and Axel, 1991; Gaillard et al., 2004). OR genes are small (~1000bp), intronless (Young and 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, *Homo sapiens*, and mouse, *Mus musculus*, respectively) (Mombaerts, 2004; Niimura and Nei, 2006; Niimura and Nei, 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 and Nei, 2006; Niimura and Nei, 2007). The total number of OR genes in a genome may reflect how many different scents can be detected and distinguished (Niimura and 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 and 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 to other primates (Rouquier et al., 2000; Gilad et al., 2004).

OR genes have been studied extensively in fish and mammals (Niimura and Nei, 2006). In 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 and Nevitt, 2004) (for reviews, see Roper, 1999; Hagelin, 2006; Hagelin and Jones, 2007).

To date, avian OR gene sequence data has 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 and Nei, 2005; Lagerstrom et al., 2006; but see Eriksson et al., 2008). An analysis of a draft red jungle fowl genomic sequence (Build 1.1; released in February 2004) reported that (i) the OR gene repertoire consisted of approximately 550 members, (ii) the predicted proportion of potentially functional OR genes was approximately 15% and (iii) the majority of the red jungle fowl OR genes clustered within a single, large clade, denoted group-γ-c (Niimura and Nei, 2005). The group-γ-c clade appears to have expanded in size after separation of the avian and mammalian lineages (Niimura and Nei, 2005) and represents an expansion of OR genes similar to the human OR5U1 and OR5BF1 genes (International Chicken Genome Sequencing Consortium, 2004; Lagerstrom et al., 2006). Note that, because the red jungle fowl genomic sequence analysed was of draft status, the estimated number and proportion of potentially functional OR genes should be considered as underestimates (Niimura and Nei, 2005). Indeed, other studies estimated the potentially functional OR gene repertoire of the Build 1.1 draft red jungle fowl genomic sequence to be either 229 (Lagerstrom 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 red jungle fowl 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 and 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 and 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.

MATERIALS AND METHODS

Amplification and sequencing of OR genes

Blood samples were suspended in Queens 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 template in subsequent amplification reactions. In total, ten primers were designed to anneal to evolutionarily conserved coding sequences corresponding to the transmembrane domain (TM) 3 (forward primers) and TM7 (reverse primers) 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 amongst 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'– ATCTGYAARCCIYTICAYTA –3' and 5'– RTTGCIATYTGYAARCCYCTRCACTA -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 µl containing , 2.0 mM Mg^{2+}, dNTPs (0.1 mM); primers (0.8 µM); 1 U *Taq* DNA polymerase (MBI Fermentas, St. Leon Rot, Germany) and genomic DNA (100 ng) template with thermocycling parameters: 94°C/ 2 minutes; 94°C/30 seconds, 37°C/30 seconds, ramping from 37°C to 72°C at 0.2°C/ second, 72°C/60 seconds, 5 cycles; 94°C/30 seconds, 45°C/30 seconds, 72°C/60 seconds, 30 cycles; 72°C/7 minutes; 4°C/hold.

Amplification products were separated through 2% agarose gels (Nusieve GTG agarose, BioWhittaker Molecular Applications, Rockland, U.S.A.), bands were excised and purified (QIAquick Gel Extraction kit, Qiagen) before ligation into a T-tailed cloning vector (pGemT-easy, Promega, Madison, U.S.A.). Note that PCR products were not pooled for ligation and transformation. Plasmids having inserts were purified from transformed DH5α 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).

Sequence analysis and phylogenetic tree construction

We obtained on average 150 ± 11 SEM (range = 98-206) sequences per species. Electropherograms were visually inspected, edited and low-quality sequences discarded. PCR primer sequences were deleted and sequences sharing ≥ 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 query in a BLAST search in the NCBI's non-redundant database. Sequences that did not return an established vertebrate OR sequence as a 'best hit' were removed from further analyses. Sequences were shifted into the correct reading frame using a custom-written PERL script. Due 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 OR partial 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 OR sequences of established classification (Niimura and Nei, 2005). Note that OR genes amplified with 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 OR partial coding sequences and the corresponding primer combinations used is shown in supplementary Table 2.1. We assigned a sequence as 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 pseudogene (Gilad et al., 2004). In nine cases, copies of the same clone 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 of the amplified coding region or mutations in promoter regions will not be detected (Gilad et al., 2004). To determine how many potentially 'functional' OR coding sequences from

the experiments are indeed pseudogenes (because of mutations outside of the amplified region), we conducted a search for OR genes in the second draft of the red jungle fowl genome (Build 2.1, released in May 2006). The red jungle fowl OR sequences identified with the PCR based method were then compared against the set of OR genes identified by the database search using a BLAST-approach. In addition, we compared the red jungle fowl sequences based on the degenerate PCR approach with Niimura and Nei's (2005) dataset, which was based on the first draft of the red jungle fowl genome.

A General Linear Mixed Model (GLMM) was used to compare the proportion of potentially functional OR genes between the γ-c and the non-γ-c clade (Venables and 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 random factor and clade as predictor variable.

Clustal X 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 (Kumar et al., 2004; http://www.megasoftware.net/). The reliability of the phylogenetic tree was evaluated with 1000 bootstrap repeats.

Estimation of OR repertoire size

A nonparametric estimation technique applying the concept of 'sample coverage' (Chao and 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 (ACE1) 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 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 supplementary Table 2.2. Note that the estimated total number of OR genes might be an underestimate of the true value (Bunge and Fitzpatrick, 1993).

Phylogenetically independent contrasts

To control for phylogenetic non-independence, we calculated phylogenetically independent contrasts (PIC; Felsenstein, 1985) using the PDAP:PDTREE module of Mesquite (Midford et al., 2005; Maddison and Maddison, 2006). The topology of the tree and branch lengths were obtained by using genetic distances derived from DNA-DNA hybridization studies (Sibley and Ahlquist, 1991; see supplementary Figure 2.1). Since we could not estimate the number of OR genes from the black coucal (see above), we obtained seven contrasts from eight species.

Supplementary Information

OR sequences generated in this study were deposited to GenBank (http://www.ncbi.nih.gov/Genbank/) with accession numbers **EF426863-EF427345** (see supplementary Table 2.1).

RESULTS

Proportion of potentially functional OR genes

We amplified 46 distinct partial OR coding sequences from red jungle fowl genomic DNA (Table 2.1; supplementary Table 2.1). 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 $\pm$ SEM; 53.5 $\pm$ 4.2) partial OR coding sequences from a further eight species representing six additional avian orders (Table 2.1; supplementary Table 2.1). The estimated proportion of potentially functional OR genes was consistently high in all taxa (mean $\pm$ SEM; 83.7% $\pm$ 2.3%) despite the wide phylogenetic distribution and diverse ecological niches of the taxa examined (Table 2.1). The estimated proportion of potentially functional OR genes was not statistically significantly different between the large γ-c OR clade (mean $\pm$ SEM; 80.8% $\pm$ 3.9 %) and the non-γ-c OR clade (mean $\pm$ SEM; 85.7% $\pm$ 2.7%) (GLMM, $t_{1,8}$ = 0.34, p = 0.74).

Comparison of data based on degenerate PCR and genome search

Eighteen of the 46 sequences that we amplified using the degenerate PCR method were identical ($\geq$98.5% nucleotide identity) to OR genes identified from the red jungle fowl genome search (Build 2.1, released in May 2006). The other 28 sequences were on average 94.9 $\pm$ 0.5% identical to OR genes identified from the red jungle fowl genome search. Because the large majority of the other sequences (27 out of 28) mapped to 'chrUn_random' regions in the red jungle fowl genome and because the Build 2.1 genome draft still contains many sequence gaps, we assume that we amplified many OR coding sequences that are not yet known.

A direct comparison of the results from the degenerate PCR and from the genome search shows that two coding sequences that 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 functional OR genes in the red jungle fowl genome by 11%. The comparison of the red jungle fowl sequence with Niimura and Nei's (2005) dataset yielded similar results (details not shown).

Table 2.1
Summary of data on relative olfactory bulb size (OBR), the OR gene repertoire (number of OR genes amplified, number of pseudogenes amplified, estimated percentage of potentially functional OR genes, estimated total OR gene numbers) and ecological variables for nine avian species.

Common name	Scientific name	Order	OBR[1]	Number of OR genes amplified[5]	Number of pseudogenes amplified[5]	% of potentially functional OR genes	Estimated total number of OR genes	Ecological variables[6] (Habitat/diet/activity pattern)
Canary	*Serinus canaria*	Passeriformes	6.0 [2]	55 (19/36/0)	11 (1/10/0)	80	166	Scrub / granivorous / diurnal
Galah	*Eolophus roseicapillus*	Psittaciformes	8.0 [2,4]	26 (17/9/0)	7 (3/4/0)	73.1	107	Grassland and agricultural areas / herbivorous / diurnal
Blue tit	*Cyanistes caeruleus*	Passeriformes	9.7 [2]	62 (18/43/1)	7 (6/1/0)	88.7	218	Forest / omnivorous /diurnal
Red jungle fowl	*Gallus gallus*	Galliformes	14.2 [2]	46 (11/35/0)	2 (0/2/0)	95.7	638	Forest / omnivorous / diurnal
Mallard	*Anas platyrhynchos*	Anseriformes	19.4 [2]	61 (14/47/1)	9 (0/9/0)	85.2	430	Wetlands / omnivorous / diurnal
Black coucal	*Centropus grillii*	Cuculiformes	19.5 [2,4]	61 (9/52/0)	8 (0/8/0)	86.9	__	Grassland / insectivorous / diurnal
Kakapo	*Strigops habroptilus*	Psittaciformes	30.2 [3]	56 (21/35/0)	10 (2/8/0)	82.1	667	Forest / herbivorous / nocturnal
Brown kiwi	*Apteryx australis*	Apterygiformes	34.0 [2]	68 (29/39/0)	16 (8/8/0)	76.5	600	Forest and scrub / insectivorous/ nocturnal
Snow petrel	*Pagodroma nivea*	Procellariiformes	37.0 [2]	47 (21/26/0)	7 (3/4/0)	85.1	212	Marine / planktonivorous and piscivorous / diurnal

[1] the olfactory bulb ratio (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 (Bang and Cobb, 1968).
[2] from Bang and Cobb (1968)
[3] from Hagelin (2004)
[4] mean OBR of this order
[5] Numbers in brackets refer to non-γ-c OR genes, γ-c OR genes, and group-α OR genes, respectively.
[5] [6] from references del Hoyo et al. (1992); Cramp and Perrins (1994); Elliot et al. (1994); Bonadonna and Bretagnolle (2002); del Hoyo et al. (2007)

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 2.1 and supplementary Table 2.1). Our estimate for the red jungle fowl, 638 OR genes, is close to the previously reported estimate of 550 derived by Niimura and Nei (2005) from a draft red jungle fowl 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, galah) are within the range reported as typical of fish (Ngai et al., 1993), while the higher values (600 - 667; red jungle fowl, brown kiwi, 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 olfactory bulb ratio (OBR), the ratio of the greatest diameter of the olfactory bulb to the greatest diameter of the cerebral hemisphere in per cent (Bang and Cobb, 1968) (number: $r = 0.63$, $n = 8$, $P < 0.05$ (1-tailed), supplementary Figure 2.2a; proportion: $r = 0.20$, $n = 9$, $P = 0.6$, supplementary Figure 2.2b).

Phylogenetic trees derived from predicted OR protein sequences

An expanded γ-c OR clade is present in all nine avian genomes examined (Figure 2.1a). 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 2.1a). In contrast, amongst 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. A NJ phylogenetic tree based on an alignment of the 405 predicted potentially functional avian OR protein sequences identified in this study (Table 2.1) and the corresponding regions of potentially functional OR proteins identified from the jungle fowl, zebrafish and human genome sequences (Niimura and 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 2.1b).

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 1.1 draft red jungle fowl genomic sequence by Niimura and Nei (2005). Our results are consistent with those of the (International Chicken Genome Sequencing Consortium, 2004), who estimated that 93% (202 of 218) of OR genes in the γ-c OR clade are potentially functional (also based on analysis of the Build 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 red jungle fowl genome (Build 2.1, released in May 2006) (Kuryshev et al., 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 2.1 draft, many putative partial OR genes still contain sequence

Figure 2.1 (opposite page)

(a) Unrooted Neighbour-Joining (NJ) phylogenetic trees derived from alignments of predicted vertebrate OR protein sequences (TM3-TM7). (a) NJ phylogenetic tree of 483 predicted avian protein sequences derived from predicted functional OR genes from the canary (*Serinus canaria*, 44 sequences, ●) the blue tit (*Cyanistes caeruleus*, 55 sequences, ●), the galah, (*Elophus roseicapillus*, 19 sequences, ●), the kakapo (*Strigops habroptilus*, 46 sequences, ●), the black coucal (*Centropus grillii*, 53 sequences, ●), the mallard (*Anas platyrhynchos*, 52 sequences, ●), the red jungle fowl (*Gallus gallus*, 44 and 78 sequences, ● and ●), the snow petrel (*Pagodroma nivea*, 40 sequences, ●) and the brown kiwi (*Apteryx australis*, 52 sequences, ●). Red jungle fowl sequences that were obtained from Niimura and Nei (2005) (n = 78) are indicated with black dots, while red jungle fowl sequences amplified in this study are indicated with yellow dots (n = 44). Note that few class I genes, indicated within a 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 tree generated from aligned vertebrate predicted OR protein sequences: human (388 sequences, black lines), zebra fish (*Danio rerio*, 98 sequences, blue lines) and avian (483 sequences, pink lines). The human and zebra fish predicted OR protein sequences were obtained from Niimura and Nei (2005) while the avian OR sequences were from Niimura and Nei (2005) or this work. The γ-c OR clade is shaded in grey. Scale-bar indicates the number of amino acid substitutions per site.

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 red jungle fowl could be as high as 85% (Kuryshev et al., 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 and Nei, 2007). The observed differences in OR gene repertoire sizes are striking, but perhaps not too surprising for the following two reasons. Firstly, birds also show wide interspecific variation in the relative olfactory bulb size, as quantified by the OBR. For example, the OBR of the snow petrel (*Pagodroma nivea*) is twelve times larger than that of the black-capped chickadee (*Poecile atricapillus*) (Bang and Cobb, 1968). Hence, similar interspecific variation in OR gene repertoire size could be expected. Secondly, in mammals, OR gene repertoire sizes range from 606 OR genes in the macaque to 2129 OR genes in the cow (Niimura and 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 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 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 characterised 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).

(a)

(b)
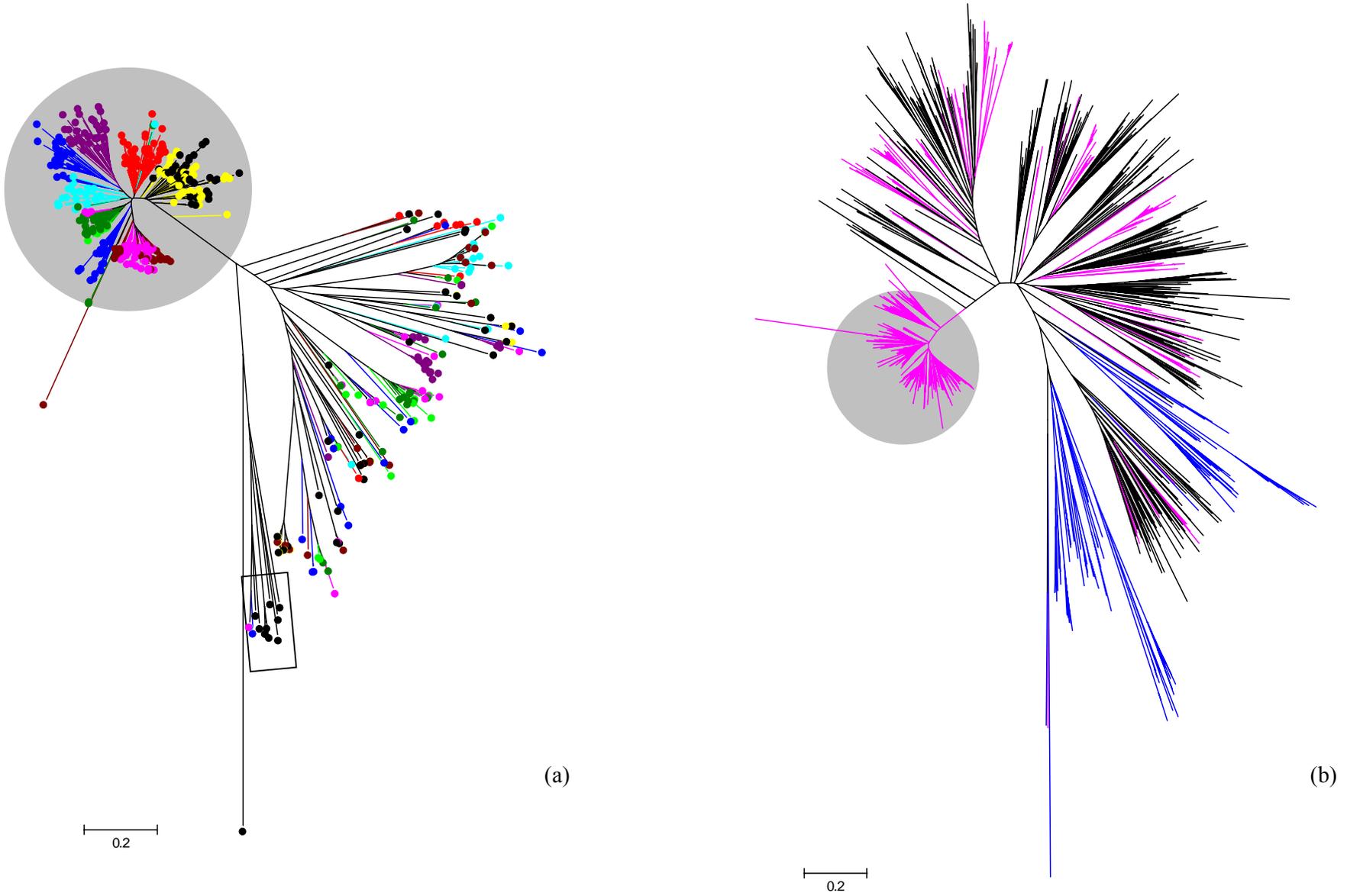
Notwithstanding its wide application, it is well recognized that this PCR-based approach has limitations and may overestimate the proportion of potentially functional OR genes, because (i) primer annealing sites may be more conserved in functional than in pseudogenes, and (ii) because mutations that 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 red jungle fowl genome by about 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 OR partial coding sequences amongst 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 ~ 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 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, 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 and 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 (*Taeniopygia guttata)*, whose genome sequence will soon become available, will be similar to that of the two passeriform genomes analysed here, ~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, amongst 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 and Nei, 2006). To account for phylogeny (Felsenstein, 1985), we based our analysis on Sibley and 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, 1-tailed). Hence, 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 and 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 and 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 and Mason, 1987). Similarly, blue tits (*Cyanistes 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 relationships 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, γ-c OR clade expansion may be a characteristic of all bird genomes. Two lines of evidence indicate that the γ-c OR clade expansion did not occur before divergence of the avian lineage. Firstly, 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 (Steiger, unpublished data). Secondly, we did not detect any group-γ-c OR genes in database searches of a draft reptilian genomic sequence (Anolis lizard, *Anolis carolinensis*) (Kuryshev et al., in prep). 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.

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 localised on microchromosome 16 (Fillon et al., 1996). However, to our knowledge, OR genes have not been located nearby. Since the majority of red jungle fowl 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 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 review, see Nei and 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 and 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 jungle fowl γ-c OR genes have not yet been assigned to specific chromosomes (Build 2.1), BLAST searches have established that the 40 jungle fowl γ-c OR genes identified by Niimura and Nei (2005) are located on 22 different contigs with a total length of 1691 kb. This represents only 0.14% of the total jungle fowl genome, suggesting that 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 the 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 function, contradict the general view that avian olfactory ability is poorly developed.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Figure 2.1
Phylogeny used for calculating phylogenetic independent contrasts of (a) the number of OR genes on olfactory bulb size ratio (OBR) for eight avian species and (b) the proportion of potentially functional OR genes on olfactory bulb size ratio for nine avian species. The topology of the tree and branch lengths were obtained by using genetic distances derived from DNA-DNA hybridization studies (Sibley and Ahlquist, 1991). Nodes are indicated with numbers.
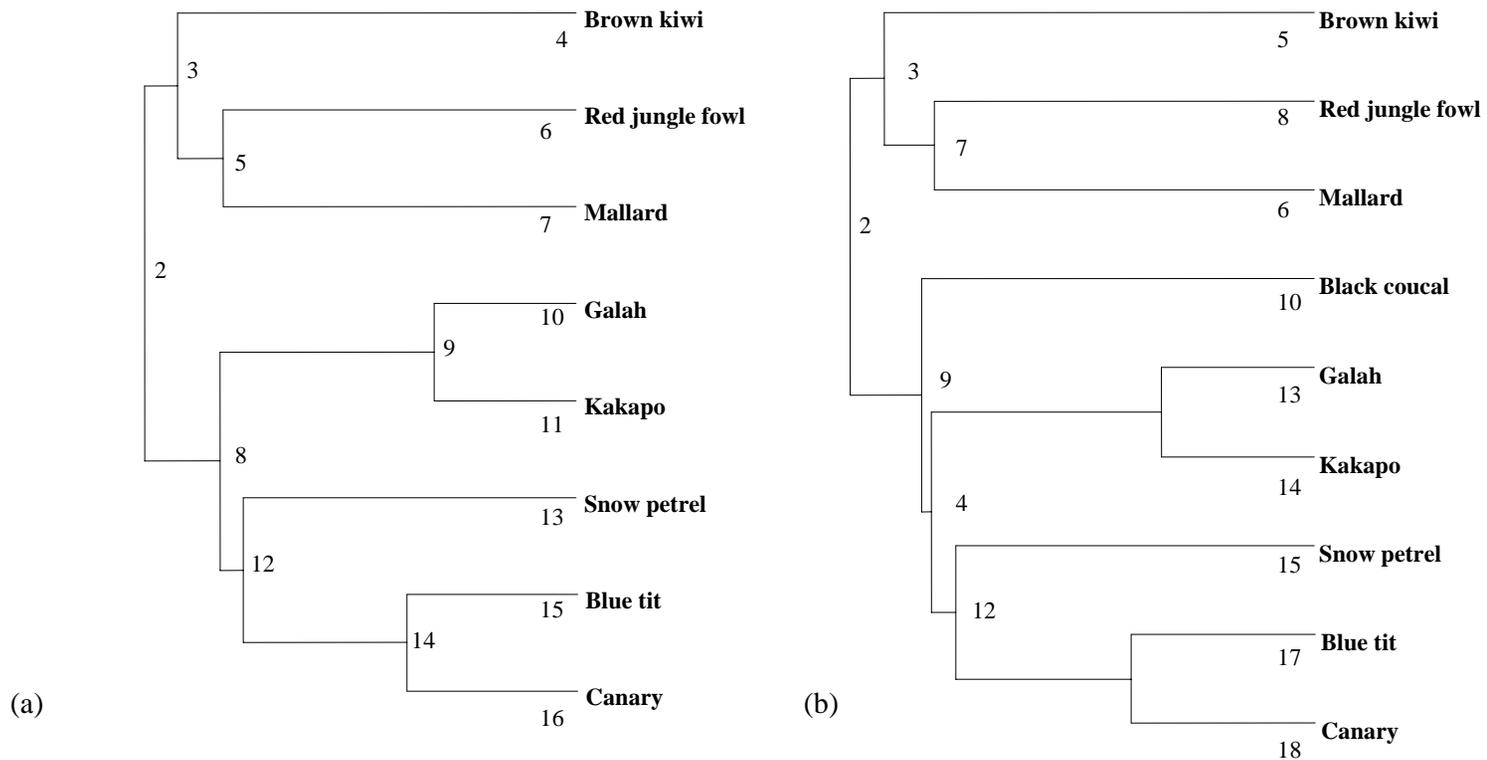
Figure 2.2
Phylogenetic independent contrasts (PIC). (a) The relationship between phylogenetic independent contrasts of the number of OR genes on olfactory bulb size ratio for eight avian species. Numbers refer to the nodes in supplementary Figure 2.1a. (b) The relationship between phylogenetic independent contrasts of the proportion of potentially functional OR genes on olfactory bulb size ratio for nine avian species. Numbers refer to the nodes in supplementary Figure 2.1b.

(a)
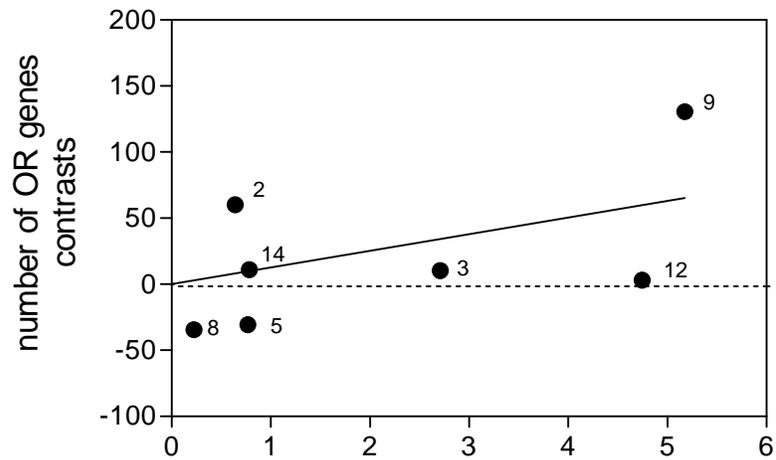
(b)
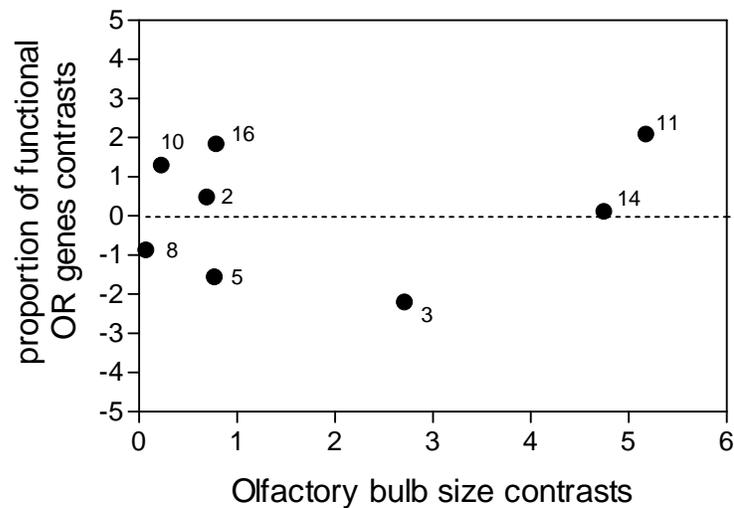
TABLE 2.1
Summary of avian olfactory receptor genes.

Common name	OR gene	Potentially functional / Pseudogene	Copies	Primer pair	GenBank accession number
Canary	ScOR1	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426925
Canary	ScOR2	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426926
Canary	ScOR3	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426927
Canary	ScOR4	Functional	4	ICKPLHY / NPFIYS(F/L)	EF426928
Canary	ScOR5	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426929
Canary	ScOR6	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF426930
Canary	ScOR7	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF426931
Canary	ScOR8	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426932
Canary	ScOR9	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426933
Canary	ScOR10	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426934
Canary	ScOR11	Functional	3	ICKPLHY / NPFIYS(F/L)	EF426935
Canary	ScOR12	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426936
Canary	ScOR13	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF426937
Canary	ScOR14	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426938
Canary	ScOR15	Functional	3	ICKPLHY / NPFIYS(F/L)	EF426939
Canary	ScOR16	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426940
Canary	ScOR17	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426941
Canary	ScOR18	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426942
Canary	ScOR19	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426943
Canary	ScOR20	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426944
Canary	ScOR21	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426945
Canary	ScOR22	Functional	3	VAICKPLHY / NPFIYS(F/L)	EF426946
Canary	ScOR23	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426947
Canary	ScOR24	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426948
Canary	ScOR25	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426949
Canary	ScOR26	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426950
Canary	ScOR27	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426951
Canary	ScOR28	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426952
Canary	ScOR29	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426953
Canary	ScOR30	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426954
Canary	ScOR31	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426955
Canary	ScOR32	Functional	5	MAYDRY / PMLNPLIY	EF426956
Canary	ScOR33	Functional	2	MAYDRY / PMLNPLIY	EF426957
Canary	ScOR34	Functional	1	MAYDRY / PMLNPLIY	EF426958
Canary	ScOR35	Functional	1	MAYDRY / PMLNPLIY	EF426959
Canary	ScOR36	Functional	9	MAYDRY / PMLNPLIY	EF426960
Canary	ScOR37	Functional	4	MAYDRY / NPFIYS(F/L)	EF426961
Canary	ScOR38	Functional	2	MAYDRY / NPFIYS(F/L)	EF426962
Canary	ScOR39	Pseudogene	1	MAYDRY / NPFIYS(F/L)	EF426963
Canary	ScOR40	Pseudogene	1	MAYDRY / NPFIYS(F/L)	EF426964
Canary	ScOR41	Functional	1	MAYDRY / NPFIYS(F/L)	EF426965
Canary	ScOR42	Functional	3	MAYDRY / NPFIYS(F/L)	EF426966
Canary	ScOR43	Pseudogene	1	MAYDRY / PM(L/F)NP	EF426967
Canary	ScOR44	Functional	1	MAYDRY / PM(L/F)NP	EF426968

Canary	ScOR45	Functional	1	MAYDRY / PM(L/F)NP	EF426969
Canary	ScOR46	Functional	1	MAYDRY / PM(L/F)NP	EF426970
Canary	ScOR47	Functional	1	MAYDRY / PM(L/F)NP	EF426971
Canary	ScOR48	Functional	4	AMAYDRY / PMLNPLIY	EF426972
Canary	ScOR49	Functional	1	AMAYDRY / PMLNPLIY	EF426973
Canary	ScOR50	Functional	1	AMAYDRY / PMLNPLIY	EF426974
Canary	ScOR51	Functional	2	AMAYDRY / PMLNPLIY	EF426975
Canary	ScOR52	Functional	1	AMAYDRY / NPFIYS(F/L)	EF426976
Canary	ScOR53	Functional	2	AMAYDRY / PM(L/F)NP	EF426977
Canary	ScOR54	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF426978
Canary	ScOR55	Functional	7	MAYDRY(V/L)AIC / PMLNPLIY	EF426979
Galah	ErOR1	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427041
Galah	ErOR2	Pseudogene	3	ICKPLHY / NPFIYS(F/L)	EF427042
Galah	ErOR3	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427043
Galah	ErOR4	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427044
Galah	ErOR5	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427045
Galah	ErOR6	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427046
Galah	ErOR7	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427047
Galah	ErOR8	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427048
Galah	ErOR9	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427049
Galah	ErOR10	Functional	2	MAYDRY / PM(L/F)NP	EF427050
Galah	ErOR11	Functional	10	MAYDRY / PM(L/F)NP	EF427051
Galah	ErOR12	Functional	2	MAYDRY / PM(L/F)NP	EF427052
Galah	ErOR13	Functional	1	MAYDRY / NPFIYS(F/L)	EF427053
Galah	ErOR14	Functional	1	MAYDRY / NPFIYS(F/L)	EF427054
Galah	ErOR15	Functional	3	MAYDRY / PMLNPLIY	EF427055
Galah	ErOR16	Functional	1	MAYDRY / PMLNPLIY	EF427056
Galah	ErOR17	Functional	3	AMAYDRY / NPFIYS(F/L)	EF427057
Galah	ErOR18	Functional	2	AMAYDRY / PM(L/F)NP	EF427058
Galah	ErOR19	Pseudogene	1	AMAYDRY / NPFIYS(F/L)	EF427059
Galah	ErOR20	Functional	2	AMAYDRY / NPFIYS(F/L)	EF427060
Galah	ErOR21	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427061
Galah	ErOR22	Functional	1	AMAYDRY / PMLNPLIY	EF427062
Galah	ErOR23	Functional	1	MAYDRY / PM(L/F)NP	EF427063
Galah	ErOR24	Pseudogene	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427064
Galah	ErOR25	Functional	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427065
Galah	ErOR26	Pseudogene	2	MAYDRY / PM(L/F)NP	EF427066
Blue tit	CcOR1	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426863
Blue tit	CcOR2	Functional	7	VAICKPLHY / NPFIYS(F/L)	EF426864
Blue tit	CcOR3	Functional	4	VAICKPLHY / NPFIYS(F/L)	EF426865
Blue tit	CcOR4	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426866
Blue tit	CcOR5	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426867
Blue tit	CcOR6	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426868
Blue tit	CcOR7	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426869
Blue tit	CcOR8	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF426870
Blue tit	CcOR9	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426871
Blue tit	CcOR10	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF426872
Blue tit	CcOR11	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426873
Blue tit	CcOR12	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426874
Blue tit	CcOR13	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426875

Blue tit	CcOR14	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426876
Blue tit	CcOR15	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426877
Blue tit	CcOR16	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426878
Blue tit	CcOR17	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426879
Blue tit	CcOR18	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426880
Blue tit	CcOR19	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426881
Blue tit	CcOR20	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426882
Blue tit	CcOR21	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426883
Blue tit	CcOR22	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426884
Blue tit	CcOR23	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426885
Blue tit	CcOR24	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426886
Blue tit	CcOR25	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426887
Blue tit	CcOR26	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426888
Blue tit	CcOR27	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426889
Blue tit	CcOR28	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426890
Blue tit	CcOR29	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426891
Blue tit	CcOR30	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426892
Blue tit	CcOR31	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426893
Blue tit	CcOR32	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426894
Blue tit	CcOR33	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426895
Blue tit	CcOR34	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426896
Blue tit	CcOR35	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426897
Blue tit	CcOR36	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426898
Blue tit	CcOR37	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426899
Blue tit	CcOR38	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426900
Blue tit	CcOR39	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426901
Blue tit	CcOR40	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF426902
Blue tit	CcOR41	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF426903
Blue tit	CcOR42	Pseudogene	1	MAYDRY / PMLNPLIY	EF426904
Blue tit	CcOR43	Pseudogene	1	MAYDRY / PMLNPLIY	EF426905
Blue tit	CcOR44	Functional	4	MAYDRY / PM(L/F)NP	EF426906
Blue tit	CcOR45	Pseudogene	8	MAYDRY / PM(L/F)NP	EF426907
Blue tit	CcOR46	Pseudogene	3	MAYDRY / PM(L/F)NP	EF426908
Blue tit	CcOR47	Functional	1	AMAYDRY / NPFIYS(F/L)	EF426909
Blue tit	CcOR48	Pseudogene	2	AMAYDRY / NPFIYS(F/L)	EF426910
Blue tit	CcOR49	Functional	1	AMAYDRY / PM(L/F)NP	EF426911
Blue tit	CcOR50	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426912
Blue tit	CcOR51	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426913
Blue tit	CcOR52	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426914
Blue tit	CcOR53	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426915
Blue tit	CcOR54	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426916
Blue tit	CcOR55	Functional	1	AMAYDRY / NPFIYS(F/L)	EF426917
Blue tit	CcOR56	Functional	2	AMAYDRY / NPFIYS(F/L)	EF426918
Blue tit	CcOR57	Functional	1	AMAYDRY / PM(L/F)NP	EF426919
Blue tit	CcOR58	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426920
Blue tit	CcOR59	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426921
Blue tit	CcOR60	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426922
Blue tit	CcOR61	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF426923
Blue tit	CcOR62	Functional	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF426924
Jungle fowl	GgOR1	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427067

Jungle fowl	GgOR2	Functional	3	ICKPLHY / NPFIYS(F/L)	EF427068
Jungle fowl	GgOR3	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427069
Jungle fowl	GgOR4	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427070
Jungle fowl	GgOR5	Functional	4	ICKPLHY / NPFIYS(F/L)	EF427071
Jungle fowl	GgOR6	Functional	4	ICKPLHY / NPFIYS(F/L)	EF427072
Jungle fowl	GgOR7	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427073
Jungle fowl	GgOR8	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427074
Jungle fowl	GgOR9	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427075
Jungle fowl	GgOR10	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427076
Jungle fowl	GgOR11	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427077
Jungle fowl	GgOR12	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427078
Jungle fowl	GgOR13	Functional	3	ICKPLHY / NPFIYS(F/L)	EF427079
Jungle fowl	GgOR14	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427080
Jungle fowl	GgOR15	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427081
Jungle fowl	GgOR16	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427082
Jungle fowl	GgOR17	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427083
Jungle fowl	GgOR18	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427084
Jungle fowl	GgOR19	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427085
Jungle fowl	GgOR20	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427086
Jungle fowl	GgOR21	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427087
Jungle fowl	GgOR22	Functional	3	VAICKPLHY / NPFIYS(F/L)	EF427088
Jungle fowl	GgOR23	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427089
Jungle fowl	GgOR24	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427090
Jungle fowl	GgOR25	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427091
Jungle fowl	GgOR26	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427092
Jungle fowl	GgOR27	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427093
Jungle fowl	GgOR28	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427094
Jungle fowl	GgOR29	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427095
Jungle fowl	GgOR30	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427096
Jungle fowl	GgOR31	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427097
Jungle fowl	GgOR32	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427098
Jungle fowl	GgOR33	Functional	4	MAYDRY / PMLNPLIY	EF427099
Jungle fowl	GgOR34	Functional	31	AMAYDRY / NPFIYS(F/L)	EF427100
Jungle fowl	GgOR35	Functional	13	AMAYDRY / PM(L/F)NP	EF427101
Jungle fowl	GgOR36	Functional	6	AMAYDRY / PMLNPLIY	EF427102
Jungle fowl	GgOR37	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427103
Jungle fowl	GgOR38	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427104
Jungle fowl	GgOR39	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427105
Jungle fowl	GgOR40	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427106
Jungle fowl	GgOR41	Functional	5	MAYDRY / PMLNPLIY	EF427107
Jungle fowl	GgOR42	Functional	3	MAYDRY / PMLNPLIY	EF427108
Jungle fowl	GgOR43	Functional	1	AMAYDRY / PMLNPLIY	EF427109
Jungle fowl	GgOR44	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427110
Jungle fowl	GgOR45	Functional	6	AMAYDRY / PM(L/F)NP	EF427111
Jungle fowl	GgOR46	Functional	4	MAYDRY / PMLNPLIY	EF427112
Mallard	ApOR1	Functional	4	ICKPLHY / NPFIYS(F/L)	EF427237
Mallard	ApOR2	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF427238
Mallard	ApOR3	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427239
Mallard	ApOR4	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427240
Mallard	ApOR5	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427241

Mallard	ApOR6	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427242
Mallard	ApOR7	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427243
Mallard	ApOR8	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427244
Mallard	ApOR9	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427245
Mallard	ApOR10	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427246
Mallard	ApOR11	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427247
Mallard	ApOR12	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427248
Mallard	ApOR13	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427249
Mallard	ApOR14	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427250
Mallard	ApOR15	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427251
Mallard	ApOR16	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427252
Mallard	ApOR17	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427253
Mallard	ApOR18	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427254
Mallard	ApOR19	Functional	3	VAICKPLHY / NPFIYS(F/L)	EF427255
Mallard	ApOR20	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427256
Mallard	ApOR21	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427257
Mallard	ApOR22	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427258
Mallard	ApOR23	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427259
Mallard	ApOR24	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427260
Mallard	ApOR25	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427261
Mallard	ApOR26	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427262
Mallard	ApOR27	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427263
Mallard	ApOR28	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427264
Mallard	ApOR29	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427265
Mallard	ApOR30	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427266
Mallard	ApOR31	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427267
Mallard	ApOR32	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427268
Mallard	ApOR33	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427269
Mallard	ApOR34	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427270
Mallard	ApOR35	Functional	5	AMAYDRY / NPFIYS(F/L)	EF427271
Mallard	ApOR36	Functional	1	MAYDRY / NPFIYS(F/L)	EF427272
Mallard	ApOR37	Pseudogene	1	MAYDRY / NPFIYS(F/L)	EF427273
Mallard	ApOR38	Functional	1	MAYDRY / NPFIYS(F/L)	EF427274
Mallard	ApOR39	Functional	1	MAYDRY / NPFIYS(F/L)	EF427275
Mallard	ApOR40	Functional	1	MAYDRY / NPFIYS(F/L)	EF427276
Mallard	ApOR41	Functional	2	MAYDRY / NPFIYS(F/L)	EF427277
Mallard	ApOR42	Functional	4	MAYDRY / PM(L/F)NP	EF427278
Mallard	ApOR43	Functional	4	AMAYDRY / PMLNPLIY	EF427279
Mallard	ApOR44	Functional	1	AMAYDRY / PMLNPLIY	EF427280
Mallard	ApOR45	Functional	15	AMAYDRY / PMLNPLIY	EF427281
Mallard	ApOR46	Functional	5	AMAYDRY / PMLNPLIY	EF427282
Mallard	ApOR47	Functional	3	AMAYDRY / PMLNPLIY	EF427283
Mallard	ApOR48	Functional	2	AMAYDRY / PMLNPLIY	EF427284
Mallard	ApOR49	Pseudogene	1	AMAYDRY / NPFIYS(F/L)	EF427285
Mallard	ApOR50	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427286
Mallard	ApOR51	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427287
Mallard	ApOR52	Functional	1	AMAYDRY / PM(L/F)NP	EF427288
Mallard	ApOR53	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427289
Mallard	ApOR54	Functional	5	AMAYDRY / NPFIYS(F/L)	EF427290
Mallard	ApOR55	Functional	1	MAYDRY / PMLNPLIY	EF427291

Mallard	ApOR56	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427292
Mallard	ApOR57	Pseudogene	2	AMAYDRY / NPFIYS(F/L)	EF427293
Mallard	ApOR58	Functional	1	MAYDRY / NPFIYS(F/L)	EF427294
Mallard	ApOR59	Functional	1	MAYDRY / NPFIYS(F/L)	EF427295
Mallard	ApOR60	Functional	5	MAYDRY / NPFIYS(F/L)	EF427296
Mallard	ApOR61	Functional	1	AMAYDRY / PMLNPLIY	EF427297
Black coucal	CgOR1	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426980
Black coucal	CgOR2	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426981
Black coucal	CgOR3	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426982
Black coucal	CgOR4	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426983
Black coucal	CgOR5	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426984
Black coucal	CgOR6	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426985
Black coucal	CgOR7	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426986
Black coucal	CgOR8	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426987
Black coucal	CgOR9	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426988
Black coucal	CgOR10	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426989
Black coucal	CgOR11	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426990
Black coucal	CgOR12	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426991
Black coucal	CgOR13	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426992
Black coucal	CgOR14	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426993
Black coucal	CgOR15	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426994
Black coucal	CgOR16	Functional	2	ICKPLHY / NPFIYS(F/L)	EF426995
Black coucal	CgOR17	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF426996
Black coucal	CgOR18	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426997
Black coucal	CgOR19	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426998
Black coucal	CgOR20	Functional	1	ICKPLHY / NPFIYS(F/L)	EF426999
Black coucal	CgOR21	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427000
Black coucal	CgOR22	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427001
Black coucal	CgOR23	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427002
Black coucal	CgOR24	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427003
Black coucal	CgOR25	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427004
Black coucal	CgOR26	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427005
Black coucal	CgOR27	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427006
Black coucal	CgOR28	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427007
Black coucal	CgOR29	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427008
Black coucal	CgOR30	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427009
Black coucal	CgOR31	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427010
Black coucal	CgOR32	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427011
Black coucal	CgOR33	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427012
Black coucal	CgOR34	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427013
Black coucal	CgOR35	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427014
Black coucal	CgOR36	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427015
Black coucal	CgOR37	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427016
Black coucal	CgOR38	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427017
Black coucal	CgOR39	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427018
Black coucal	CgOR40	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427019
Black coucal	CgOR41	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427020
Black coucal	CgOR42	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427021
Black coucal	CgOR43	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427022
Black coucal	CgOR44	Functional	28	AMAYDRY / PM(L/F)NP	EF427023

Black coucal	CgOR45	Functional	7	MAYDRY(V/L)AIC / PMLNPLIY	EF427024
Black coucal	CgOR46	Functional	9	MAYDRY(V/L)AIC / PMLNPLIY	EF427025
Black coucal	CgOR47	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427026
Black coucal	CgOR48	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427027
Black coucal	CgOR49	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427028
Black coucal	CgOR50	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427029
Black coucal	CgOR51	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427030
Black coucal	CgOR52	Pseudogene	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427031
Black coucal	CgOR53	Pseudogene	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427032
Black coucal	CgOR54	Functional	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427033
Black coucal	CgOR55	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427034
Black coucal	CgOR56	Functional	1	MAYDRY / PMLNPLIY	EF427035
Black coucal	CgOR57	Functional	1	MAYDRY / NPFIYS(F/L)	EF427036
Black coucal	CgOR58	Functional	6	MAYDRY / NPFIYS(F/L)	EF427037
Black coucal	CgOR59	Functional	3	AMAYDRY / PMLNPLIY	EF427038
Black coucal	CgOR60	Functional	1	AMAYDRY / PMLNPLIY	EF427039
Black coucal	CgOR61	Functional	2	AMAYDRY / NPFIYS(F/L)	EF427040
Snow petrel	PnOR1	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427298
Snow petrel	PnOR2	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427299
Snow petrel	PnOR3	Pseudogene	4	ICKPLHY / NPFIYS(F/L)	EF427300
Snow petrel	PnOR4	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427301
Snow petrel	PnOR5	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427302
Snow petrel	PnOR6	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427303
Snow petrel	PnOR7	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427304
Snow petrel	PnOR8	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427305
Snow petrel	PnOR9	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427306
Snow petrel	PnOR10	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF427307
Snow petrel	PnOR11	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427308
Snow petrel	PnOR12	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427309
Snow petrel	PnOR13	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427310
Snow petrel	PnOR14	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427311
Snow petrel	PnOR15	Functional	4	VAICKPLHY / NPFIYS(F/L)	EF427312
Snow petrel	PnOR16	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427313
Snow petrel	PnOR17	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427314
Snow petrel	PnOR18	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427315
Snow petrel	PnOR19	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427316
Snow petrel	PnOR20	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427317
Snow petrel	PnOR21	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427318
Snow petrel	PnOR22	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427319
Snow petrel	PnOR23	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427320
Snow petrel	PnOR24	Functional	2	VAICKPLHY / NPFIYS(F/L)	EF427321
Snow petrel	PnOR25	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427322
Snow petrel	PnOR26	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427323
Snow petrel	PnOR27	Functional	11	AMAYDRY / PMLNPLIY	EF427324
Snow petrel	PnOR28	Functional	3	MAYDRY(V/L)AIC / PM(L/F)NP	EF427325
Snow petrel	PnOR29	Functional	1	MAYDRY / PMLNPLIY	EF427326
Snow petrel	PnOR30	Functional	4	MAYDRY / NPFIYS(F/L)	EF427327
Snow petrel	PnOR31	Functional	1	MAYDRY / PMLNPLIY	EF427328
Snow petrel	PnOR32	Functional	3	AMAYDRY / PM(L/F)NP	EF427329
Snow petrel	PnOR33	Functional	1	MAYDRY / PMLNPLIY	EF427330

Snow petrel	PnOR34	Functional	6	MAYDRY / NPFIYS(F/L)	EF427331
Snow petrel	PnOR36	Pseudogene	1	AMAYDRY / NPFIYS(F/L)	EF427333
Snow petrel	PnOR37	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427334
Snow petrel	PnOR38	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427335
Snow petrel	PnOR39	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427336
Snow petrel	PnOR40	Functional	1	MAYDRY / PMLNPLIY	EF427337
Snow petrel	PnOR41	Functional	4	MAYDRY / NPFIYS(F/L)	EF427338
Snow petrel	PnOR42	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427339
Snow petrel	PnOR43	Functional	1	MAYDRY / NPFIYS(F/L)	EF427340
Snow petrel	PnOR44	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427341
Snow petrel	PnOR45	Pseudogene	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427342
Snow petrel	PnOR46	Pseudogene	2	AMAYDRY / PMLNPLIY	EF427343
Snow petrel	PnOR47	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427344
Snow petrel	PnOR48	Functional	1	MAYDRY / NPFIYS(F/L)	EF427345
Kakapo	ShOR1	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF427113
Kakapo	ShOR2	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427114
Kakapo	ShOR3	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427115
Kakapo	ShOR4	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427116
Kakapo	ShOR5	Functional	5	ICKPLHY / NPFIYS(F/L)	EF427117
Kakapo	ShOR6	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427118
Kakapo	ShOR7	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427119
Kakapo	ShOR8	Functional	3	ICKPLHY / NPFIYS(F/L)	EF427120
Kakapo	ShOR9	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427121
Kakapo	ShOR10	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427122
Kakapo	ShOR11	Functional	7	ICKPLHY / NPFIYS(F/L)	EF427123
Kakapo	ShOR12	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427124
Kakapo	ShOR13	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427125
Kakapo	ShOR14	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427126
Kakapo	ShOR15	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427127
Kakapo	ShOR16	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427128
Kakapo	ShOR17	Pseudogene	4	ICKPLHY / NPFIYS(F/L)	EF427129
Kakapo	ShOR18	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427130
Kakapo	ShOR19	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427131
Kakapo	ShOR20	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427132
Kakapo	ShOR21	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427133
Kakapo	ShOR22	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427134
Kakapo	ShOR23	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427135
Kakapo	ShOR24	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427136
Kakapo	ShOR25	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427137
Kakapo	ShOR26	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427138
Kakapo	ShOR27	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427139
Kakapo	ShOR28	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427140
Kakapo	ShOR29	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427141
Kakapo	ShOR30	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427142
Kakapo	ShOR31	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427143
Kakapo	ShOR32	Functional	13	MAYDRY / PMLNPLIY	EF427144
Kakapo	ShOR33	Functional	12	MAYDRY / NPFIYS(F/L)	EF427145
Kakapo	ShOR34	Functional	1	AMAYDRY / PMLNPLIY	EF427146
Kakapo	ShOR35	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427147
Kakapo	ShOR36	Functional	3	AMAYDRY / PM(L/F)NP	EF427148

Kakapo	ShOR37	Functional	3	MAYDRY(V/L)AIC / PMLNPLIY	EF427149
Kakapo	ShOR38	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427150
Kakapo	ShOR39	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427151
Kakapo	ShOR40	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427152
Kakapo	ShOR41	Functional	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427153
Kakapo	ShOR42	Pseudogene	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427154
Kakapo	ShOR43	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427155
Kakapo	ShOR44	Functional	4	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427156
Kakapo	ShOR45	Functional	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427157
Kakapo	ShOR46	Functional	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427158
Kakapo	ShOR47	Pseudogene	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427159
Kakapo	ShOR48	Functional	1	MAYDRY / NPFIYS(F/L)	EF427160
Kakapo	ShOR49	Pseudogene	1	MAYDRY(V/L)AIC / PM(L/F)NP	EF427161
Kakapo	ShOR50	Functional	1	MAYDRY / PMLNPLIY	EF427162
Kakapo	ShOR51	Functional	1	MAYDRY / PMLNPLIY	EF427163
Kakapo	ShOR52	Functional	1	MAYDRY / NPFIYS(F/L)	EF427164
Kakapo	ShOR53	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427165
Kakapo	ShOR54	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427166
Kakapo	ShOR55	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427167
Kakapo	ShOR56	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427168
Brown kiwi	AaOR1	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427169
Brown kiwi	AaOR2	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427170
Brown kiwi	AaOR3	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427171
Brown kiwi	AaOR4	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427172
Brown kiwi	AaOR5	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427173
Brown kiwi	AaOR6	Functional	3	ICKPLHY / NPFIYS(F/L)	EF427174
Brown kiwi	AaOR7	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427175
Brown kiwi	AaOR8	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427176
Brown kiwi	AaOR9	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427177
Brown kiwi	AaOR10	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427178
Brown kiwi	AaOR11	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427179
Brown kiwi	AaOR12	Functional	2	ICKPLHY / NPFIYS(F/L)	EF427180
Brown kiwi	AaOR13	Pseudogene	2	ICKPLHY / NPFIYS(F/L)	EF427181
Brown kiwi	AaOR14	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427182
Brown kiwi	AaOR15	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427183
Brown kiwi	AaOR16	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427184
Brown kiwi	AaOR17	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427185
Brown kiwi	AaOR18	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427186
Brown kiwi	AaOR19	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427187
Brown kiwi	AaOR20	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427188
Brown kiwi	AaOR21	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427189
Brown kiwi	AaOR22	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427190
Brown kiwi	AaOR23	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427191
Brown kiwi	AaOR24	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427192
Brown kiwi	AaOR25	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427193
Brown kiwi	AaOR26	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427194
Brown kiwi	AaOR27	Pseudogene	1	VAICKPLHY / NPFIYS(F/L)	EF427195
Brown kiwi	AaOR28	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427196
Brown kiwi	AaOR29	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427197
Brown kiwi	AaOR30	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427198

Brown kiwi	AaOR31	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427199
Brown kiwi	AaOR32	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427200
Brown kiwi	AaOR33	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427201
Brown kiwi	AaOR34	Pseudogene	1	ICKPLHY / NPFIYS(F/L)	EF427202
Brown kiwi	AaOR35	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427203
Brown kiwi	AaOR36	Functional	1	ICKPLHY / NPFIYS(F/L)	EF427204
Brown kiwi	AaOR37	Pseudogene	2	VAICKPLHY / NPFIYS(F/L)	EF427205
Brown kiwi	AaOR38	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427206
Brown kiwi	AaOR39	Functional	1	VAICKPLHY / NPFIYS(F/L)	EF427207
Brown kiwi	AaOR40	Functional	1	MAYDRY / NPFIYS(F/L)	EF427208
Brown kiwi	AaOR41	Functional	1	AMAYDRY / PMLNPLIY	EF427209
Brown kiwi	AaOR42	Pseudogene	11	AMAYDRY / PMLNPLIY	EF427210
Brown kiwi	AaOR43	Functional	3	AMAYDRY / NPFIYS(F/L)	EF427211
Brown kiwi	AaOR44	Functional	2	AMAYDRY / NPFIYS(F/L)	EF427212
Brown kiwi	AaOR45	Functional	2	AMAYDRY / NPFIYS(F/L)	EF427213
Brown kiwi	AaOR46	Functional	2	MAYDRY / NPFIYS(F/L)	EF427214
Brown kiwi	AaOR47	Functional	3	AMAYDRY / PM(L/F)NP	EF427215
Brown kiwi	AaOR48	Functional	4	AMAYDRY / PM(L/F)NP	EF427216
Brown kiwi	AaOR49	Pseudogene	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427217
Brown kiwi	AaOR50	Functional	2	MAYDRY(V/L)AIC / PMLNPLIY	EF427218
Brown kiwi	AaOR51	Pseudogene	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427219
Brown kiwi	AaOR52	Pseudogene	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427220
Brown kiwi	AaOR53	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427221
Brown kiwi	AaOR54	Functional	3	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427222
Brown kiwi	AaOR55	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427223
Brown kiwi	AaOR56	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427224
Brown kiwi	AaOR57	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427225
Brown kiwi	AaOR58	Pseudogene	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427226
Brown kiwi	AaOR59	Pseudogene	2	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427227
Brown kiwi	AaOR60	Functional	1	AMAYDRY / NPFIYS(F/L)	EF427228
Brown kiwi	AaOR61	Pseudogene	2	MAYDRY / NPFIYS(F/L)	EF427229
Brown kiwi	AaOR62	Pseudogene	1	MAYDRY / PMLNPLIY	EF427230
Brown kiwi	AaOR63	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427231
Brown kiwi	AaOR64	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427232
Brown kiwi	AaOR65	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427233
Brown kiwi	AaOR66	Functional	1	MAYDRY(V/L)AIC / NPFIYS(F/L)	EF427234
Brown kiwi	AaOR67	Functional	1	MAYDRY(V/L)AIC / PMLNPLIY	EF427235
Brown kiwi	AaOR68	Functional	1	MAYDRY / NPFIYS(F/L)	EF427236

TABLE 2.2
Abundance coverage estimators and related statistics for nine avian species. As abundance coverage estimator, ACE_1 was used (Chao and Lee, 1992).

Common name	Estimate	CI (lower)[1]	CI (upper)[1]	SEM[2]	n[3]	D[4]	C[5]	CV[6]
Canary	166.0	106.5	225.5	59.5	101	55	0.66	0.88
Galah	107.0	61.0	153.0	66.0	47	26	0.66	1.09
Blue tit	218.2	134.1	302.3	84.1	99	62	0.59	0.86
Jungle fowl	637.5	416.3	858.7	405.4	125	46	0.77	1.95
Mallard	430.3	178.3	682.3	252.0	113	61	0.63	1.32
Black coucal					115	61	0.565	2.561
Snow petrel	211.6	98.8	570.3	106.2	86	47	0.64	1.08
Kakapo	666.9	386.5	947.3	338.8	106	56	0.60	1.53
Brown kiwi	599.9	387.8	812.0	397.6	99	68	0.48	1.20

[1] CI = Confidence interval of the estimate
[2] SEM = Standard error of the estimate
[3] n = Number of clones analyzed
[4] D = Distinct number of cloned genes
[5] C = Estimation of the sample coverage
[6] CV = Estimation of the coefficient of variation

Chapter 3

Evidence for increases in olfactory receptor gene repertoires in two nocturnal bird species with well-developed olfactory ability

Silke S. Steiger, Andrew E. Fidler and Bart Kempenaers

ABSTRACT

In vertebrates, the molecular basis of the sense of smell is encoded by members of a large gene family, namely olfactory receptor (OR) genes. The number of OR genes and the proportion of functional OR genes in a genome may indicate the importance of the sense of smell for an animal. There is behavioral, physiological, and anatomical evidence that some bird species, in particular nocturnal birds, have a well developed sense of smell. Therefore, we hypothesized that birds with good olfactory abilities have evolved (i) more OR genes and (ii) more functional OR genes than closely related and presumably less 'olfaction-dependent' avian taxa. We used non-radioactive Southern hybridization and PCR using degenerate primers to investigate whether two nocturnal bird species that are known to heavily rely on olfactory cues, the brown kiwi (*Apteryx australis*) and the kakapo (*Strigops habroptilus*), have evolved a larger OR gene repertoire than their day-active, closest living relatives (emu *Dromaius novaehollandiae*, rhea *Rhea americana*, ostrich *Struthio camelus* and kaka *Nestor meridionalis*, kea *Nestor notabilis*, respectively). Here, we show that the nocturnal birds did not have a significantly higher proportion of functional OR genes. However, the estimated number of OR genes was larger in the two nocturnal birds than in their diurnal closest relatives. Finally, we show that the relative size of the olfactory bulb, a morphological indicator of olfactory ability, was positively correlated with the estimated total number of OR genes but not with the proportion of functional OR genes. In summary, our results suggest that ecological niche adaptations such as daily activity patterns may have shaped avian OR gene repertoires.

INTRODUCTION

In vertebrates, the detection of odorous chemicals in both air and water, is mediated by olfactory receptors (ORs) (Buck and Axel, 1991). ORs are members of the superfamily of seven transmembrane G-protein coupled receptors (GPCRs) and are expressed in olfactory neurons of the olfactory epithelium (Buck and Axel, 1991; for review, see Mombaerts, 2004; Niimura and Nei, 2006). OR gene coding regions are short (~ 1kb) and intronless (Buck and Axel, 1991). Both the total number and functional proportion of OR genes varies greatly amongst the genomes of the different vertebrate taxa that have been examined (Young and Trask, 2002). For example, there are about 150 OR genes in the zebra fish (*Danio rerio*) genome (Alioto and Ngai, 2005), 229 - 550 in the red jungle fowl (*Gallus gallus*) genome (International Chicken Genome Sequencing Consortium, 2004; Niimura and Nei, 2005; Lagerstrom et al., 2006) and, typically, >1000 OR genes in mammalian genomes (for review, see Gaillard et al., 2004; Ache and Young, 2005; Niimura and Nei, 2005). Amongst primates, the proportion of functional OR genes is significantly reduced in humans (~50%) when compared with other apes (~70%) (Rouquier et al., 2000; Gilad et al., 2004), a finding hypothesized to reflect decreasing behavioral reliance on the sense of smell during human evolution. More generally, it has been suggested that the total number and/or the functional proportion of OR genes in a vertebrate genome correlates with olfactory acuity at the behavioral level (Niimura and Nei, 2006).

For birds, visual and auditory cues play important roles, in behaviors as diverse as foraging, mate attraction and territory defense (Ball and Hulse, 1998; Cuthill et al., 2000). However, the significance of avian chemosensation is still debated. Increasing evidence, from both behavioral and morphological studies, suggests that olfactory ability in at least some bird species, in particular nocturnal birds, is excellent and may even be equivalent to the olfactory abilities that mammals possess (Healy and Guilford, 1990; for review, see Roper, 1999; Hagelin, 2006; Hagelin and Jones, 2007).

Concordantly with these studies the OR gene repertoire in the red jungle fowl, as estimated from of the draft genome sequence, is surprisingly large (International Chicken Genome Sequencing Consortium, 2004; Niimura and Nei, 2005; Lagerstrom et al., 2006). More recent evidence suggests that the proportion of functional OR genes is high (between 73 and 96 %) and the estimated number of OR genes is between 107 and 667 OR genes in nine distantly related avian species from seven different orders and thus, much higher than expected when working from the assumption that birds generally have a poorly developed sense of smell (Steiger et al., submitted). Interestingly, the number but not the proportion of functional OR genes was correlated with the olfactory bulb ratio (OBR) (Steiger et al, submitted), a morphological indicator of olfactory ability (Edinger, 1908).

The aim of the present study was twofold. First, whereas we previously examined the OR gene repertoires of distantly related species (Steiger et al., submitted) we now investigated the OR gene repertoires of closely related species. In particular, we investigated the OR gene repertoires of two sets of closely related species, whereby one species in each group is nocturnal and known to rely on olfactory cues. The two nocturnal bird species selected for this study were the brown kiwi (*Apteryx australis*) and the kakapo (*Strigops habroptilus*). Both species are flightless, do not have well-developed (night) vision or hearing (Martin et al., 2007, Ron J. Moorhouse, personal

communication), but do have a well-developed sense of smell and relatively large olfactory bulbs (Bang and Cobb, 1968; Hagelin, 2004). Furthermore, the brown kiwi and kakapo belong to entirely different evolutionary lineages, evolved their nocturnal behavior independently and have extant diurnal relatives, allowing comparative studies (brown kiwi relatives: emu *Dromaius novaehollandiae*, rhea *Rhea americana*, ostrich *Struthio camelus*; kakapo relatives: kaka *Nestor meridionalis*, kea *Nestor notabilis*). We expected both nocturnal birds to have evolved more OR genes than their day-active relatives. Second, we wanted to test in a new and independent dataset whether the previously reported correlation between the number of OR genes and the OBR held.

As genome sequences are not available for any avian species other than the red jungle fowl, we have used two complementary methods to estimate OR gene repertoire size. Southern hybridisation was used to provide a relative measure of the number of OR gene sequences in the avian genomes. While having the advantage of directly investigating the genome, Southern hybridisation does not provide information on the functional status of the hybridising sequences. Therefore, PCR, using degenerate primers directed at the OR gene family, and subsequent sequencing were used to estimate both the functional percentage and total number of OR genes in the avian genomes being compared. Both methods have been used previously to estimate OR gene repertoire sizes in fish (Sun et al., 1999; Irie-Kushiyama et al., 2004), amphibians (Freitag et al., 1998), birds (i.e. chicken, Leibovici et al., 1996; Nef and Nef, 1997) and mammals (Issel-Tarver and Rine, 1996; Issel-Tarver and Rine, 1997). The OR gene repertoire estimates obtained using these two methods were compared and interpreted in the light of the evolutionary history and ecological adaptations of the kiwi and kakapo.

MATERIALS AND METHODS

PCR amplification of partial olfactory receptor (OR) genes

Blood samples were stored in Queen's lysis buffer before genomic DNA was isolated using a commercial kit (DNeasy tissue kit; Qiagen, Hilden, Germany). The design of PCR primers to amplify avian OR partial coding sequences has been described in detail in Steiger et al. (submitted). Briefly, degenerate primers were annealed to evolutionarily conserved sequence motifs within transmembrane domain (TM) regions 3 and TM7 of the ORs. As a subset of the avian ORs, termed γ-c, is greatly expanded in number within avian genomes, two different sets of PCR primers were required; those targeting the more diverse non-γ-c OR genes and those targeting the more homogenous γ-c OR genes. To amplify non-γ-c OR sequences three different forward primers were used in combination with three different reverse primers (for more details, see Steiger et al., submitted). Amplifications were conducted using each forward primer in combination with each reverse primer thereby generating nine different PCR products. The conserved sequence motif within the TM3 region of the γ-c ORs differs significantly from that of the non-γ-c ORs so alternative forward primers were required for amplifying γ-c OR partial coding sequences. The two γ-c OR forward primers (for more details, see Steiger et al., submitted) were used in combination with the reverse primer 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. Amplification products were separated through 2% (w/v)

agarose gels (Nusieve GTG agarose, BioWhittaker Molecular Applications, Rockland, U.S.A.) and products of ~0.5 kb cloned and sequenced as described in Steiger et al. (submitted). Note that amplification products generated using the non-γ-c and the γ-c OR clade primers were pooled using equal volume aliquots before the ligation reaction. Therefore there were two ligations: one using the nine heterogenous non-γ-c amplicons and one using the three heterogenous γ-c amplicons.

Sequence editing and analysis

Electropherograms were visually inspected and low-quality sequences discarded. PCR primer sequences were deleted and sequences sharing ≥98.5% identity, determined using the Sequence Identity Matrix function of BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), were considered to be amplified from a single OR orthologue (Fuchs et al., 2001). To confirm that the sequences were partial OR coding sequences, each was used as a query string in BLASTX searches of the GenBank non-redundant (nr) database. Sequences were shifted into the correct reading frame using a custom-written PERL program. For each species, 50 individual clones were analyzed (25 clones derived by using primer pairs that predominantly amplified the non-γ-c OR clade and the γ-c OR clade, respectively).

Estimation of the proportion of functional OR genes and the total number of OR genes and correlation with the olfactory bulb size ratio (OBR)

We assigned a sequence as potentially functional gene if an uninterrupted coding region was found (i.e. sequence without stop codon), and as pseudogene when an interrupted coding region was found (i.e. sequence with stop codon; Gilad et al. 2004). Note that this method may overestimate the proportion of functional OR genes, because frame-shift mutations outside of the amplified region or in promoter regions will not be detected (Rouquier et al., 2000; Gilad et al., 2004). In two cases, copies of the same clone were both functional and pseudogenes and these were excluded from further analysis. Chi-square tests were used in SPSS 15.0 (SPSS, Chicago, IL, USA) to compare the proportion of functional OR genes between the nocturnal birds and their diurnal relatives.

As described previously in Steiger et al. (submitted), a nonparametric estimation technique applying the concept of 'sample coverage' (Chao and Lee, 1992) was used to estimate the total number of OR genes in each genome 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 (ACE1) 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 primer bias. 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 supplementary Table 3.1. Note that the estimated total number of OR genes might be an underestimate of the true value (Bunge and Fitzpatrick, 1993). In a previous study, we estimated the jungle fowl OR gene repertoire to consist of 638 genes (Steiger et al., submitted), which was close to the previous estimate of 550 (Niimura and Nei, 2005). This suggested that our method provides a reliable estimate of OR gene repertoire sizes in species for which full genomic sequences are not yet available.

Olfactory bulb size ratios (OBR) were taken from Bang and Cobb (1968) and Hagelin (2004) and correlated with both the estimated proportion of functional OR genes and the total number of OR genes. The OBR is the ratio of the greatest diameter of the olfactory bulb to the greatest diameter of the cerebral hemisphere in per cent and is supposed to be a morphological indicator of olfactory ability (Edinger, 1908). Note that OBR values from the kea, the kaka and the ostrich are unknown. As OBRs generally tend to be consistent within avian orders (Bang and Cobb, 1968), we used a mean value of known OBRs from 3 psittaciform and 3 palaeognaths species, respectively. Although the kakapo is known to have an exceptionally large OBR relative to other species of the same order (Bang and Cobb, 1968; Hagelin, 2004) and the kiwi has one of the largest OBR observed in birds, we included them for the calculation of the mean OBR. This leads to a more conservative test of our hypothesis, because the real OBR values for the kea, kaka and ostrich are probably smaller.

To control for phylogenetic non-independence, we calculated phylogenetically independent contrasts (Felsenstein, 1985) (PIC) using the PDAP:PDTree module of Mesquite (Midford et al., 2005; Maddison and Maddison, 2006). The topology of the tree was adapted from Cooper et al. (2001) and de Kloet and de Kloet (2005) and can be found in supplementary Figure 3.1. Branch lengths were set to 1.

Genomic Southern Hybridisation Analysis

For Southern hybridisation analysis, genomic DNA (10 µg) was digested with four restriction endonucleases (*Eco*RI, *Pst*I, *Hind*III and *Taq*I; Fermentas, St. Leon-Rot Germany) and the digestion products electrophoresed through 0.8% agarose gels before transfer to positively-charged nylon membranes (Roche, Germany; Whatman Schleicher&Schuell, Dassel, Germany) using a vacuum blotting system (VacuGeneXL; Amersham Biosciences, Freiburg, Germany) following the manufacturers protocol.

OR probes were generated from the elegant-crested tinamou (*Eudromia elegans*; family Tinamiformes, outgroup to the other ratites) and the galah (*Elophus roseicapillus*; family Cacatuidae, outgroup to the other parrots), respectively. It is important to note that the taxa used as outgroups are equally distantly related to all of the in-group taxa (Figure 3.1).

For probe generation, partial OR coding sequences were cloned into pGemT-easy and their sequences determined (see above). Plamids were digested with the restriction endonuclease EcoRI (Fermentas, St. Leon-Rot, Germany) and the inserts isolated from agarose gels (QIAquick Gel Extraction kit, Qiagen, Hilden, Germany).

(a)

(b)
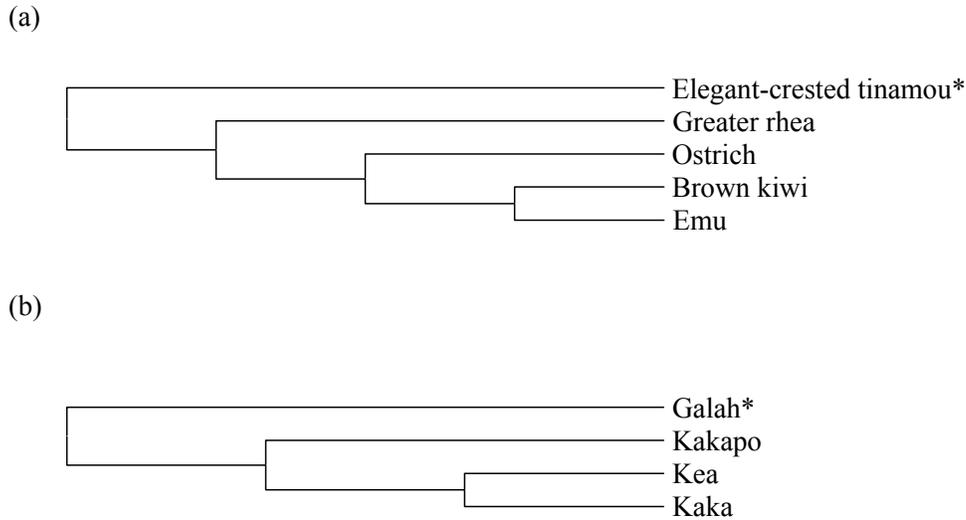
Figure 3.1
Phylogenetic relationships among avian taxa investigated in this study. (a) Palaeognath phylogenetic tree topology adapted from Cooper et al. (2001) (b) Partial pssitaciform phylogenetic tree, topology derived from de Kloet and de Kloet (2005). Outgroup taxa used for the generation of probes for Southern hybridsation are indicated by an asterisk. Branch lengths were arbitrarily set to 1.

Nucleic acid sequences of the probes were aligned with 78 known, putatively functional red jungle fowl OR receptor genes (from Niimura and Nei, 2005; sequences listed in Supporting data set 8, http://www.pnas.org/cgi/content/full/0501922102/DC1/13) using ClustalW (Thompson et al., 1994). We used the Neighbour-Joining (NJ) method and Poisson-distances to construct phylogenetic trees using the MEGA software package (version 4.0; http://www.megasoftware.net/; Tamura et al., 2007). The reliability of this phylogeny was evaluated with 1000 bootstrap repeats (Figure 3.2). Four elegant-crested tinamou and three galah OR partial coding sequences were selected as probes (denoted Tin-A – D, accession numbers **EU599489 – EU599492**, and Gal-A – C, accession numbers **EU599486 – EU599488**, respectively). Probes selected for the Southern blot hybridization were sampled from distinct regions of the OR phylogenetic tree. Their positions within the red jungle fowl OR phylogenetic tree is indicated in Figure 3.2. Sequence identity of the probes on the nucleic and amino acid level ranged between 49 - 66% and 34% - 59 %, respectively, and is thus lower than the threshold of cross-hybridisation (Lancet and Ben-Arie, 1993) (Table 3.1).

As control, a β-actin probe was generated: β-actin is a highly conserved sequence and was expected to hybridize equally with all the genomes being compared. Briefly, forward primer 5'-GAGAAATTGTGCGTGACATCA-3' and reverse primer 5'-CCTGAACCTCTCATTGCCA-3' were used to amplify a ~150 bp product of β-actin in the chicken (*Gallus gallus*) (Kubo et al., 2006). The PCR conditions were as follows: 94°C/ 5 minutes, 1 cycle; 94°C/ 30 seconds; 50°C/ 30 seconds, 72°C, 60 seconds; 30 cycles; 72°C/7 minutes. See above for plasmid digestion and insert isolation.

OR probes (and the *β*-actin probe) were labeled with digoxigenin (DIG) using a commercial kit (DIG high prime DNA Labeling and Detection Starter Kit I, Roche, Germany). Blots were hybridized with ~25ng/ml probe overnight under low (37°C) and/or high (42°C) stringency conditions.

Note that due to the limited amount of genomic DNA available from the kakapo we conducted Southern blots comparing the OR gene repertoire of the kakapo, kea and kaka only under high stringency and thus, more conservative conditions.

Washes and visualization with NBT/BCIP were done following the manufacturers protocol. Membranes were photographed by a gel documentation system (Gel Doc 2000, BIO-RAD). The brightness and the contrast of the images were optimized with Photoshop (Adobe Photoshop)**.**

Supplementary Information

The OR sequences generated in this study were deposited on GenBank (http://www.ncbi.nih.gov/Genbank/) with accession numbers **EU594675-EU594890** and **EU599486 – EU599492** (for more details, see supplementary Table 3.2).

Figure 3.2 (opposite page)
Neighbour-joining (NJ) phylogenetic tree generated from an alignment of 78 predicted red jungle fowl (*Gallus gallus)* putatively functional OR nucleotide sequences as described by Niimura and Nei (2005). Positions of sequences derived from the partial OR cDNA sequences used as probes in the Southern blotting probes are indicated: TinA-D, GalA-C. Probes Gal-C and Tin-D are clearly placed within the γ-c OR clade while probes Gal-A, B and Tin-A,B,C are placed within the non-γ-c OR clades. Bootstrap values > 80% are indicated. Scale-bar indicates the number of nucleic acid substitutions per site. Abbreviations: Tin, elegant-crested tinamou, Gal, galah; OR, olfactory receptor.
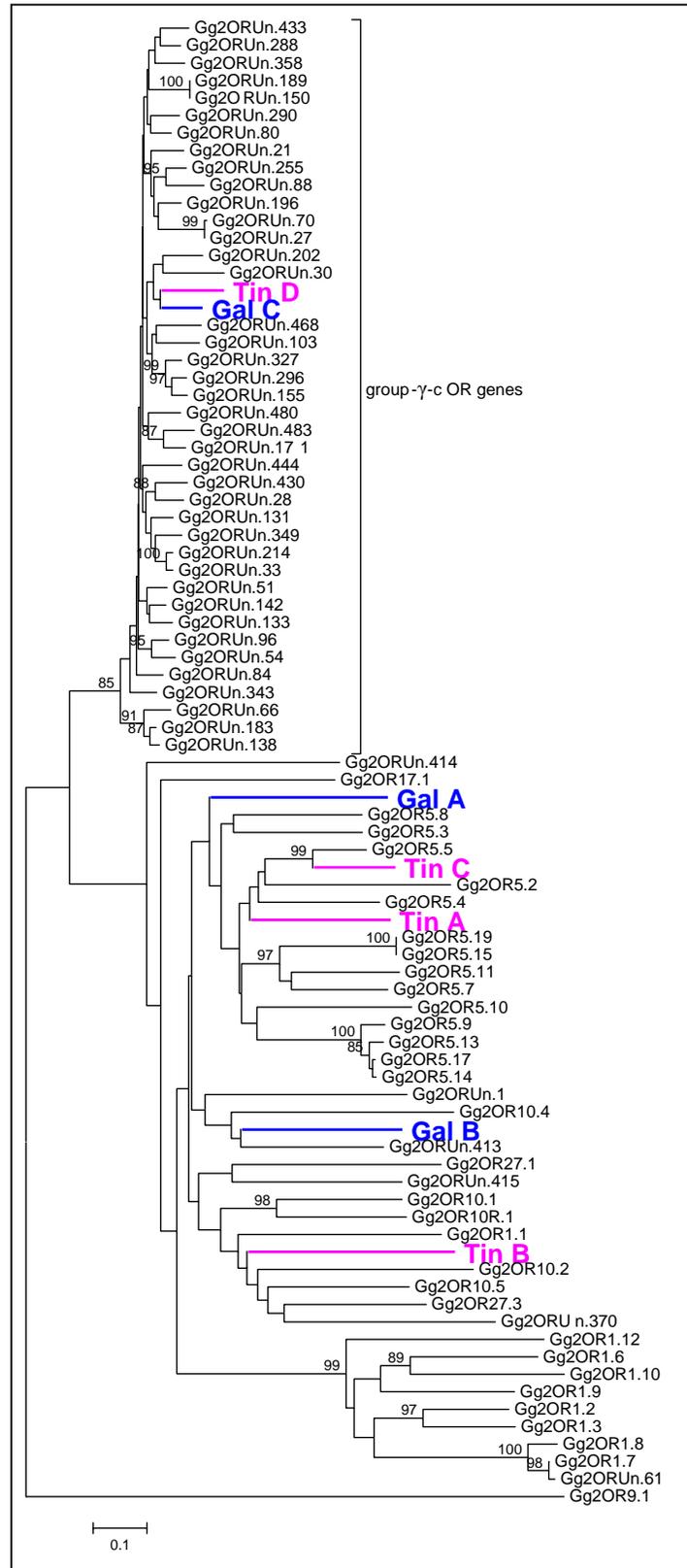
Table 3.1
Sequence identities (in %) between (a) elegant-crested tinamou and (b) galah probes on the nucleic / amino acid level. GenBank accession numbers of the probe sequences are: Tin-A = **EU599489**, Tin-B = **EU599490**, Tin-C = **EU599491**, Tin-D = **EU599492**; Gal-A = **EU599486**, Gal-B = **EU599487**, Gal-C = **EU599488**.

(a)

Probes	Tin-A	Tin-B	Tin-C	Tin-D
Tin-A		52 / 38	66 / 59	52 / 41
Tin-B	52 / 38		56 / 44	49 / 33
Tin-C	66 / 59	56 / 44		56 / 40
Tin-D	52 / 59	49 / 33	56 / 40	

(b)

Probes	Gal-A	Gal-B	Gal-C
Gal-A		56 / 42	51 / 37
Gal-B	56 / 42		49 / 34
Gal-C	51 / 37	49 / 34	

RESULTS

Estimation of the proportion of potentially functional OR genes

The total numbers of distinct OR sequences amplified with the degenerate PCR primers from the seven avian genomes varied between 22 and 42 (Table 3.2). The mean number of distinct OR sequences was similar in both taxonomic groups (mean ± SEM: paleognaths: 32 ± 4, psittaciformes: 30 ± 5). Among the paleognaths, the proportion of functional OR genes did not differ significantly (mean± SEM: 79.6% ± 1.9%; $\chi^2 = 0.7$, df =3, P=0.87). Although the mean proportion of functional OR genes was similar in the psittaciformes (79.9% ± 7.5%), the proportions differed significantly between the three species ($\chi^2 = 6.3$, df=2, P=0.04). Notably, the proportion of functional OR genes was significantly higher in the kea and kaka than in the kakapo (Table 3.2). The lower percentage of predicted functional OR genes in the kakapo was primarily caused by a large number of pseudogenes (12) amplified from the γ-c OR clade.

Estimation of OR gene clade sizes using Southern blot hybridization

(a) Analysis of OR gene numbers in palaeognath genomes

Genomic DNA from the brown kiwi, emu, greater rhea, ostrich and elegant-crested tinamou was digested with four different restriction enzymes and transferred to a filter before hybridization at low stringency (37 ºC) with four elegant-crested tinamou OR gene probes, Tin-A - D.

The Tin-A probe yielded more bands in the emu than in the brown kiwi, the ostrich or the rhea (Figure 3a), indicating that this subfamily is expanded in the emu. The Tin-B subfamily was relatively small: the rhea and emu showed only 1 and 2 bands, whereas in the kiwi between two and seven bands can be seen (Figure 3.3b). The ostrich seems to entirely lack this OR gene subfamily: only one very faint band in the lane containing ostrich genomic DNA digested with PstI can be seen. The Tin-C subfamily yielded more bands in the kiwi than in any of the other species (Figure 3.3c). Note that although sequence identity between Tin-C and Tin-A is relatively high, banding patterns are clearly different, indicating that these probes reflect different subfamilies despite clustering closely together on the phylogenetic tree (Figure 3.1).

The Tin-D subfamily (group-γ-c OR) was largest in all species studied (Figure 3.3d), which reflects its placement within the large γ-c OR clade (Figure 3.2). Although quantification is difficult due to the very large number of bands, the intensity of the 'smear' in the brown kiwi lanes is stronger than in the other taxa (Figure 3.1d). This difference is most prominent in the EcoRI and HindIII digested samples where the intensity of the hybridization to the brown kiwi DNA is clearly stronger than that to the ostrich, emu, rhea and even tinamou. This suggests that multiple brown kiwi sequences hybridise with the Tin-D probe and are flanked by EcoRI and HindIII sites. It was difficult to discern changes in gene number due to the large size of this subfamily.

Although one subfamily seems to be greatly expanded in the emu, these data overall suggest a larger OR gene repertoire in the brown kiwi.

Common name	Scientific name	Order	Family	Activity pattern	OBR (in %)	Number of pseudogenes amplified	Number of distinct OR sequences amplified	Proportion of potentially functional OR genes (%)	Estimated number of OR genes
Paleognath comparison									
Brown kiwi	*Apteryx australis*	Apterygiformes	Apterygidae	nocturnal	34[a]	9	42	78.6	478
Emu	*Dromaius novaehollandiae*	Casuariiformes	Dromaiidae	diurnal	26.3[a]	6	31	80.6	109
Greater Rhea	*Rhea americana*	Rheiformes	Rheidae	diurnal	19[a]	7	28	75.0	66
Ostrich	*Struthio camelus*	Struthioniformes	Struthionidae	diurnal	26.4[b]	4	25	84	58
Elegant-crested tinamou[e]	*Eudromia elegans*	Tinamiformes	Tinamidae	diurnal					
Psittaciform comparison									
Kakapo	*Strigops habroptilus*	Psittaciformes	Psittacidae[f]	nocturnal	30.2[c]	12	38	68.4	312
Kaka	*Nestor meridionalis*	Psittaciformes	Psittacidae[f]	diurnal	15.4[d]	5	22	77.3	55
Kea	*Nestor notabilis*	Psittaciformes	Psittacidae[f]	diurnal/ crepuscular	15.4[d]	2	30	93.3	102
Galah[e]	*Eolophus roseicapillus*	Psittaciformes	Cacatuidae	diurnal					

[a] data from Bang (1968)
[b] mean OBR of Paleognathae (Ratites included: Brown kiwi, Emu, Greater Rhea)
[c] from Hagelin (2004)
[d] mean OBR of Psittaciformes (Psittaciformes included: Kakapo; Rose-ringed Parakeet, *Psittacula krameri*; Budgeriar, *Melopsittacus undulatus*)
[e] used as outgroup
[f] Kakapo have also been classified as the sole member of the family Strigopidae; Keas and Kakas have been placed in their own family Nestoridae (Christidis and Boles, 2008)

Table 3.2 (opposite page)
Overview about the species studied and their OR gene repertoires. Activity pattern, olfactory bulb ratio (OBR; measured as the ratio of the greatest diameter of the bulb to the greatest diameter of the cerebral hemisphere in per cent; Bang, 1968), number of pseudogenes amplified, predicted proportion of potentially functional OR genes and estimated number of OR genes for seven bird species. Note that for each species, 50 individual clones were analyzed (see Methods).

(b) Analysis of OR gene numbers in psittaciform genomes

Genomic DNA from the kakapo, kea, kaka and galah was digested with four different restriction enzymes and transferred to a filter before hybridization at high stringency (42 ºC) with three galah OR gene probes, Gal-A, -B and -C.

In all four pssitaciform taxa, multiple fragments of varying intensity were detected with the Gal-A and -B probes (Figure 3.4a-b). The fragment patterns for the kea and the kaka are strikingly similar for both probes with all four restriction enzymes (Figure 3.4a-b), reflecting the close taxonomic relationship between these two parrot species. In contrast, the fragment patterns of the kakapo, are quite distinct from those of the kea and kaka, particularly in the *Taq*I and *Pst*I digests (Figure 3.4a-b). The kakapo lanes contain more bands or, in the case of the PstI/Gal-B combination, a much more intense band than the kea and kaka or, indeed, the galah. In general, the variable intensity is presumed to reflect varying levels of sequence homology with the probes. However, the fact that the intensity of hybridization to the kakapo DNA is much stronger than that to the galah (the species from which the probe was derived), suggests that multiple kakapo sequences hybridised with the probe.

The Gal-C probe generated the largest number of bands when hybridized to all four pssitaciform genomes (Figure 3.4c), which reflects its placement within the large γ-c OR clade (Figure 3.1). Similar to the results obtained in the palaeognath genome comparison, quantification is difficult due to the very large number of bands. Note that the intensity of the 'smear' in the kakapo lanes is stronger than in the other taxa (Figure 3.4c). This difference is most prominent in the *Pst*I digested samples where the intensity of the hybridization to the kakapo DNA is clearly stronger than to the kea, kaka and even the galah. This again suggests that multiple kakapo sequences hybridized with the Gal-C probe and are flanked by *Pst*I sites.

Note that the observed differences in hybridization intensities observed in both species groups cannot be attributed to differences in the amounts of DNA loaded into each lane, because both ethidium bromide staining of the gels and hybridization with a *β*-actin probe showed no differences in signal intensities between lanes (data not shown).

Estimation of the total number of OR genes

The estimated total number of OR genes in each genome varied between 55 and 478. The estimated OR gene repertoires of the nocturnal species (kakapo and kiwi) were 5 to 8 times larger than those found in their diurnal closest relatives (Table 3.2).

Figure 3.3 (opposite page)
Southern hybridization of restriction enzyme digested palaeognathae genomic DNAs. Genomic DNA isolated from five palaeognath taxa (ostrich, emu, brown kiwi, greater rhea, elegant-crested tinamous) was digested with four different restriction enzymes and used for Southern hybridization with four DIG-labeled probes generated from galah partial OR coding sequences: (a) probe Tin-A, (b) probe Tin-B, (c) probe Tin-C and (d) probe Tin-D. GenBank accession numbers of the probe sequences are: Tin-A = **EU599489**, Tin-B = **EU599490**, Tin-C = **EU599491**, Tin-D = **EU599492**. Low stringency hybridization conditions (37ºC hybridization temperature) were used for all results shown. Abbreviations: T, *Taq*I; P, *Pst*I; E, *Eco*RI; H, *Hind*III; DIG, digoxigenin; Tin, elegant-crested tinamou. Approximate positions of the size standards (kb) are indicated.

Correlation of OBR with the proportion of functional OR and the estimated total number of OR genes

The estimated total number of OR genes, but not the predicted proportion of functional OR genes correlated positively with OBR (proportion: PIC, r= -0.52, n=6, P=0.23, number: PIC, r =0.8, n=6, P=0.03). The relationships between phylogenetic independent contrasts of (A) the estimated number of OR genes in avian genomes on olfactory bulb ratio (OBR) and (B) the proportion of functional OR genes on OBR can be found in supplementary Figure 3.2.

DISCUSSION

In this study, two molecular techniques (PCR and Southern blotting) were used to compare the numbers of OR genes and the proportion of functional OR genes across avian genomes. Both techniques have their limitations. First, PCR using degenerate primers may overestimate the number of functional OR genes, as these presumably have stronger conservation of primer annealing sites than do pseudogenes (Gilad et al., 2004). Secondly, only half of the open region frame was amplified and sequenced and therefore, mutations occurring in the remaining N- and C-termed coding regions are not detected. Finally, due to unpredictable primer bias, some OR genes may amplify preferentially. Thus, the ratios of OR partial coding sequences amongst the amplification products may not represent a random sample of the OR repertoires in the template DNA (Rouquier et al., 2000; Gilad et al., 2004). The limitation of the Southern Blot method is that it only estimates the number of OR genes in a given subfamily and does not provide information about the functionality of these genes. However, it should be emphasized that both methods compensate for each techniques limitations.

The majority of OR coding sequences amplified from all of the avian genomes in this study did not have stop codons, suggesting that they derive from functional genes. The degree of interspecific variation in the proportion of functional genes appears to be small. This is consistent with previous work, which showed a high proportion of functional OR
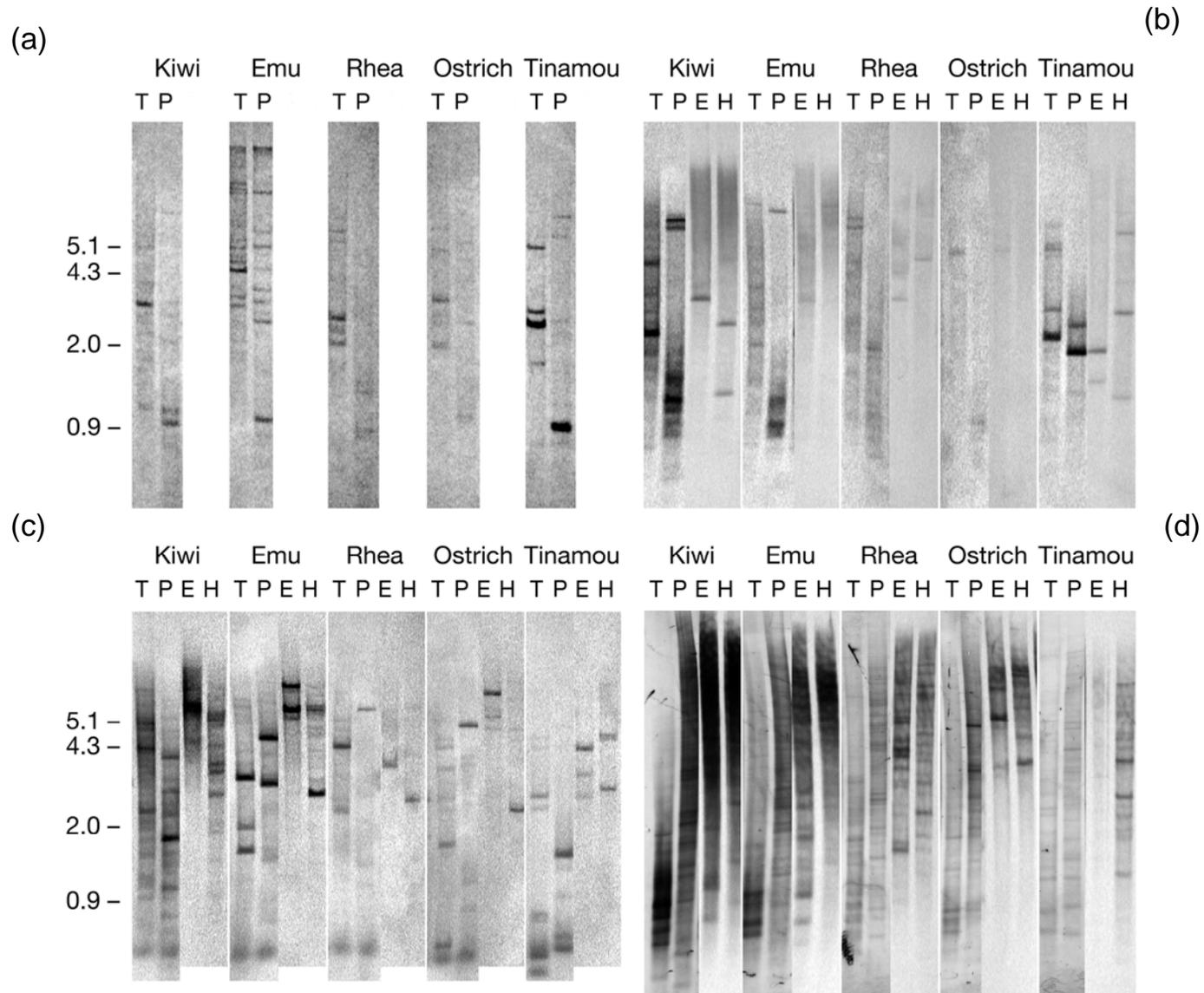
Figure 3.4 (opposite page)
Southern hybridization of restriction enzyme digested psittacciform genomic DNAs. Genomic DNA isolated from four psittaciform taxa (kaka, kea, kakapo and galah) was digested with four different restriction enzymes and used for Southern hybridization with three DIG-labeled probes generated from galah partial OR coding sequences: (a) probe Gal-A, (b) probe Gal-B and (c) probe Gal-C. GenBank accession numbers of the probe sequences are: Gal-A = **EU599486**, Gal-B = **EU599487**, Gal-C = **EU599488**. High stringency hybridization conditions were used for all results shown (42ºC hybridization temperature). Abbreviations: T, *Taq*I; P, *Pst*I; E, *Eco*RI; H, *Hind*III; DIG, digoxigenin, Gal, galah. Approximate positions of the size standards (kb) are indicated.

genes in nine bird genomes from seven different orders, including the kiwi and the kakapo (Steiger et al., submitted). The estimate of the proportion of functional genes for the kakapo was higher in this previous study (82.1% versus 68.4% in this study), possibly due to the slightly different method used (the PCR products were pooled in this study and thus, there were two ligation reactions, whereas previously they were not pooled and there were nine ligation reactions). Notably, this difference was not significant ($\chi^2 = 2.3$, df=1, P=0.12).

If the proportion of functional OR genes indicated olfactory ability, one may expect (nocturnal birds) with good olfactory abilities to encode a higher proportion of functional OR genes than diurnal birds. Contrary to this hypothesis, the (i) estimated proportion of functional OR genes were lower for the nocturnal kakapo than for the kea and the kaka and (ii) did not differ between the kiwi and its diurnal relatives. However, the estimated total number of OR genes was 5 to 8 times larger in the two nocturnal species than in their diurnal relatives, even though our estimate of the kakapo OR gene repertoire was smaller than in the previous study (Steiger et al., submitted). Thus, it may be the total number of OR genes rather than the proportion of functional OR genes that is most closely linked with olfaction in birds (Steiger et al., submitted).

The results of the Southern Blot analyses generally agree with the estimates of OR gene numbers based on degenerate PCR. Both nocturnal species showed generally more or stronger bands than their diurnal closest relatives. This was particularly clear in the expanded γ-c OR gene clade, indicating that the two nocturnal species do have a greater total number of OR genes.

Olfactory receptors evolve dynamically (Niimura and Nei, 2006) via duplication and/or gene conversion in a process that has been called 'birth and death evolution' (Nei et al., 1997).Why would it be advantageous for a nocturnal bird to have evolved more (and maybe more diverse) olfactory receptor genes? It has been suggested that the more OR genes there are in the genome the finer the discrimination amongst odour molecules is (Alioto and Ngai, 2005).Thus, differences in the size of the OR gene repertoire may cause different odour sensitivities among birds. A wide receptor repertoire is also likely to be required for binding/detection of many, structurally-diverse, 'odourous' compounds (Young et al., 2002). Thus, a nocturnal bird that has evolved more OR genes might be able to smell (i) more diverse odorants and (ii) odorants at a much lower concentration than a diurnal bird that lives in the same habitat. The large number of estimated OR genes in the two nocturnal species may contribute to the ability of these species to locate food at night via olfactory cues (Wenzel, 1968; Hagelin, 2004). Due to the absence of terrestrial

(a)

(b)

(c)
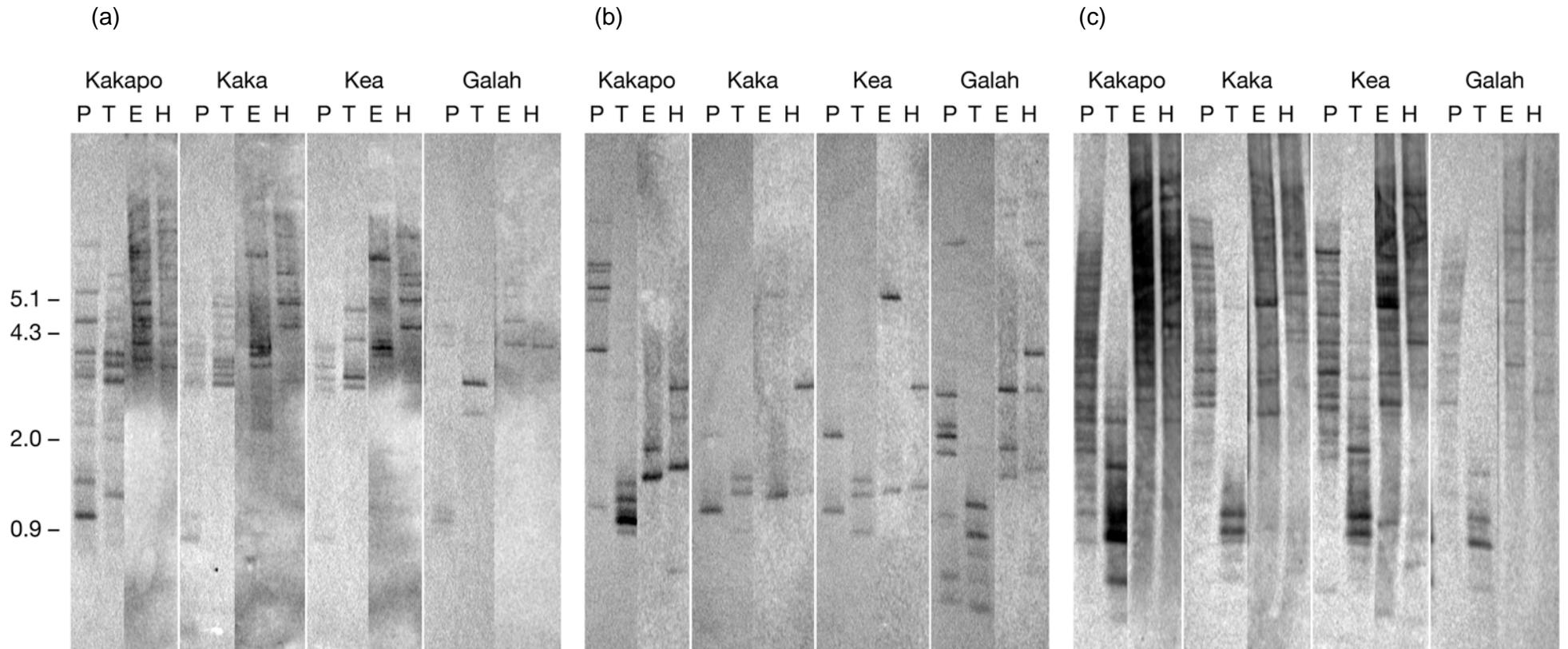
mammalian predators in New Zealand (Wilson, 2004), it is perhaps not surprising that these nocturnal birds have adapted to the echological niche usually occupied by mammals. This includes the development of such mammalian-like characteristic as reliance upon olfactory information (Martin et al., 2007).

We further found that the relative size of the olfactory bulb correlated with the estimated total number of OR genes, but not with the predicted proportion of functional OR genes, again confirming earlier work (Steiger et al., submitted) Interestingly, the olfactory anatomy is also remarkably well developed in procellariiform seabirds (petrels, albatrosses and shearwaters) and olfactory cues such as krill-related odours or odours associated with phytoplankton play an important role in foraging behaviour (Bang and Cobb, 1968; Nevitt, 2000). So far, procellariiform OR genes have received little attention (Steiger et al., submitted). Therefore, future studies could investigate whether the reliance on olfactory cues in seabirds is reflected in the OR gene repertoires. For example, it would be interesting to determine whether burrowing petrels that return to their nest at night have evolved a larger OR gene repertoire than surface-nesting petrels that rather rely on visual cues to recognize their nest (Bonadonna and Bretagnolle, 2002).

Besides enhancement of the olfactory system, nocturnal birds can compensate for the reduced effectiveness of vision at night by an enhanced sense of vision (e.g. owls; Koenig and Becking, 1999), or by increasing other capacities such as hearing (owls, (Koenig and Becking, 1999)) or echolocation (oilbirds and swiftlets) (Medway, 1959; Konishi and Knudsen, 1979). Further work could also address whether nocturnal birds that invested in visual perception (e.g. owls, Strigiformes) have evolved smaller olfactory receptor repertoires, which would suggest a trade-off between investment in vision versus olfaction.

In summary, out data indicate that the OR gene repertoire is larger in two nocturnal bird species than in closely related, but diurnal bird species. Our results confirm and extend previous behavioral studies suggesting that nocturnal bird species have a well-developed sense of smell. Our results strongly suggest that ecological niche adaptations such as daily activity patterns may have shaped avian OR gene repertoires. This is remarkable, as this supports the notion that birds – similar to mammals - may heavily rely on their sense of smell.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Figure 3.1
Phylogeny used in calculating phylogenetic independent contrasts (PICs) of the number and functional proportion of OR genes on olfactory bulb ratio (OBR). The tree topology follows Cooper et al. (2001) for the ratites and de Kloet and de Kloet (2005) for the parrots. All nodes are indicated with numbers and are referenced in additional file 2. Branch lengths were set to 1.
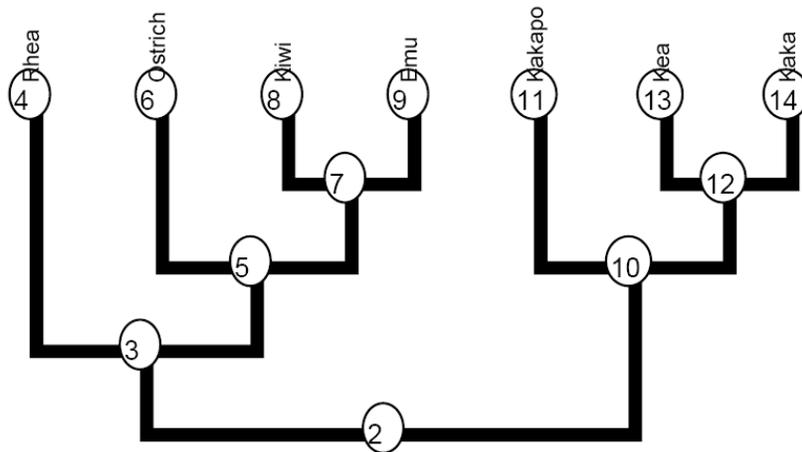
Figure 3.2
The relationships between phylogenetic independent contrasts of (a) the estimated number of
OR genes in avian genomes on olfactory bulb ratio (OBR) and (b) the proportion of functional
OR genes on OBR. The relationship in (a) is significantly positive while that in (b) is not.
Numbers refer to the nodes in supplementary Figure 1.

(a)

(b)

Olfactory bulb size contrasts
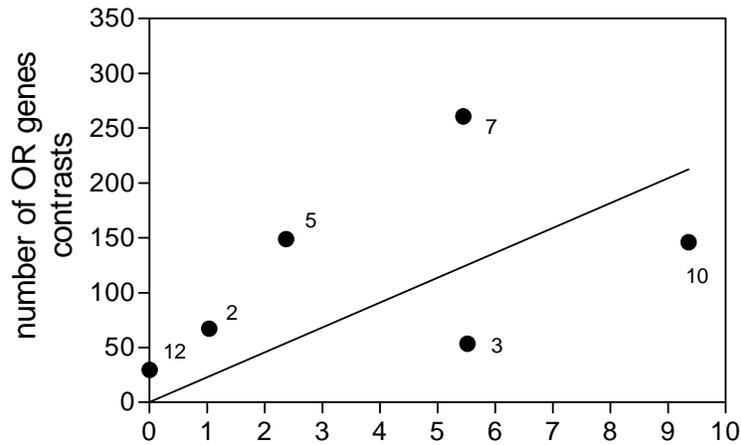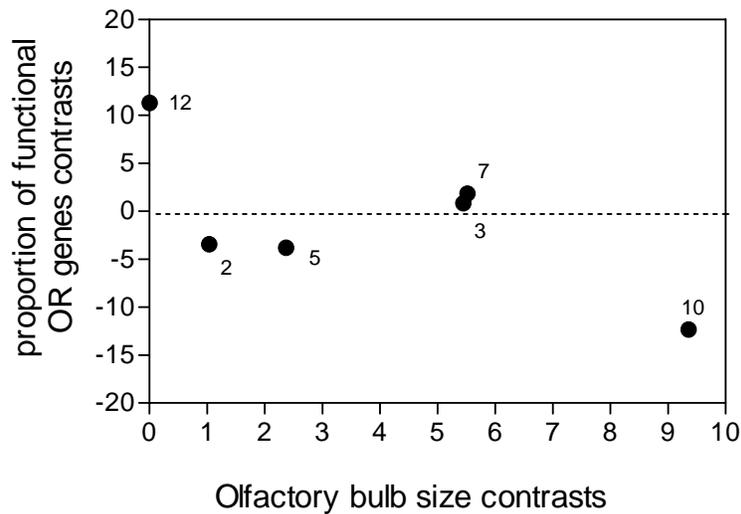
Table 3.1
Abundance coverage estimators and related statistics for seven avian species. The abundance coverage estimator ACE_1 was used, calculated as described in Chao and Lee, 1992.

Common name	Estimator	Estimate	CI95_low[2]	CI95_high[2]	Est_se[3]	N[4]	D[5]	C[6]	CV[7]
Paleognath comparison									
Brown kiwi	ACE_1[1]	477.8	156.0	1707.4	336.6	50	42	0.260	0.920
Emu	ACE_1[1]	108.6	56.2	274.6	49.2	50	32	0.540	0.740
Ostrich	ACE_1[1]	58.2	34.4	142.6	23.9	50	25	0.700	0.738
Rhea	ACE_1[1]	66.4	39.5	156.2	26.0	50	28	0.660	0.699
Psittaciform comparison									
Kaka	ACE_1[1]	55.4	30.9	154.9	26.5	50	23	0.790	0.975
Kakapo	ACE_1[1]	312.3	122.1	932.0	181.6	50	38	0.360	0.946
Kea	ACE_1[1]	102.1	52.4	262.6	47.2	50	30	0.580	0.804

[1] ACE_1 = Abundance coverage estimator for heterogeneous samples
[2] CI_95low, CI_95high = Confidence interval of the estimate (low & high boundary)
[3] Est_se = Standard error of the estimate
[4] n = Number of plasmids sequenced
[5] D = Number of distinct OR partial coding sequences
[6] C = Estimation of the sample coverage
[7] CV = Estimation of coefficient of variation

TABLE 3.2
Summary of (a) paleognath and (b) psittaciform partial olfactory receptor (OR) sequences
generated in this study. Copies indicate how often a partial OR sequenced was amplified.

(a)

Common name	OR gene	Potentially functional/Pseudogene	Copies	GenBank accession number
Emu	EmuOR1	Pseudogene	1	EU594675
Emu	EmuOR2	Functional	1	EU594686
Emu	EmuOR3	Functional	1	EU594697
Emu	EmuOR4	Functional	1	EU594700
Emu	EmuOR5	Functional	3	EU594701
Emu	EmuOR6	Functional	4	EU594702
Emu	EmuOR7	Functional	5	EU594703
Emu	EmuOR8	Functional	1	EU594704
Emu	EmuOR9	Pseudogene	1	EU594705
Emu	EmuOR10	Functional	1	EU594676
Emu	EmuOR11	Pseudogene	2	EU594677
Emu	EmuOR12	Pseudogene	1	EU594678
Emu	EmuOR13	Functional	2	EU594679
Emu	EmuOR14	Functional	1	EU594680
Emu	EmuOR15	Functional	1	EU594681
Emu	EmuOR16	Functional	1	EU594682
Emu	EmuOR17	Functional	1	EU594683
Emu	EmuOR18	Functional	3	EU594684
Emu	EmuOR19	Functional	1	EU594685
Emu	EmuOR20	Pseudogene	2	EU594687
Emu	EmuOR21	Functional	1	EU594688
Emu	EmuOR22	Functional	1	EU594689
Emu	EmuOR23	Functional	1	EU594690
Emu	EmuOR24	Functional	1	EU594691
Emu	EmuOR25	Functional	1	EU594692
Emu	EmuOR26	Functional	1	EU594693
Emu	EmuOR27	Functional	1	EU594694
Emu	EmuOR28	Pseudogene	1	EU594695
Emu	EmuOR29	Functional	1	EU594696
Emu	EmuOR30	Functional	4	EU594698
Emu	EmuOR31	Functional	2	EU594699
Kiwi	KiwiOR1	Pseudogene	1	EU594849
Kiwi	KiwiOR2	Functional	1	EU594860
Kiwi	KiwiOR3	Functional	1	EU594871
Kiwi	KiwiOR4	Pseudogene	1	EU594882
Kiwi	KiwiOR5	Functional	1	EU594886
Kiwi	KiwiOR6	Functional	1	EU594887
Kiwi	KiwiOR7	Functional	1	EU594888
Kiwi	KiwiOR8	Functional	2	EU594889
Kiwi	KiwiOR9	Functional	1	EU594890
Kiwi	KiwiOR10	Functional	1	EU594850
Kiwi	KiwiOR11	Peudogene	1	EU594851
Kiwi	KiwiOR12	Functional	1	EU594852
Kiwi	KiwiOR13	Functional	1	EU594853
Kiwi	KiwiOR14	Pseudogene	1	EU594854
Kiwi	KiwiOR15	Functional	1	EU594855
Kiwi	KiwiOR16	Functional	1	EU594856
Kiwi	KiwiOR17	Functional	5	EU594857
Kiwi	KiwiOR18	Functional	1	EU594858

Kiwi	KiwiOR19	Pseudogene	1	EU594859
Kiwi	KiwiOR20	Functional	1	EU594861
Kiwi	KiwiOR21	Functional	1	EU594862
Kiwi	KiwiOR22	Functional	2	EU594863
Kiwi	KiwiOR23	Functional	1	EU594864
Kiwi	KiwiOR24	Functional	1	EU594865
Kiwi	KiwiOR25	Pseudogene	1	EU594866
Kiwi	KiwiOR26	Functional	1	EU594867
Kiwi	KiwiOR27	Pseudogene	1	EU594868
Kiwi	KiwiOR28	Functional	1	EU594869
Kiwi	KiwiOR29	Functional	1	EU594870
Kiwi	KiwiOR30	Functional	1	EU594872
Kiwi	KiwiOR31	Functional	1	EU594873
Kiwi	KiwiOR32	Functional	2	EU594874
Kiwi	KiwiOR33	Functional	1	EU594875
Kiwi	KiwiOR34	Functional	1	EU594876
Kiwi	KiwiOR35	Functional	1	EU594877
Kiwi	KiwiOR36	Functional	1	EU594878
Kiwi	KiwiOR37	Functional	1	EU594879
Kiwi	KiwiOR38	Functional	1	EU594880
Kiwi	KiwiOR39	Functional	1	EU594881
Kiwi	KiwiOR40	Pseudogene	1	EU594883
Kiwi	KiwiOR41	Pseudogene	1	EU594884
Kiwi	KiwiOR42	Functional	2	EU594885
Ostrich	OstrichOR1	Functional	1	EU594796
Ostrich	OstrichOR2	Pseudogene	5	EU594807
Ostrich	OstrichOR3	Functional	6	EU594814
Ostrich	OstrichOR4	Functional	5	EU594815
Ostrich	OstrichOR5	Functional	1	EU594816
Ostrich	OstrichOR6	Functional	1	EU594817
Ostrich	OstrichOR7	Functional	1	EU594818
Ostrich	OstrichOR8	Functional	1	EU594819
Ostrich	OstrichOR9	Pseudogene	2	EU594820
Ostrich	OstrichOR10	Functional	1	EU594797
Ostrich	OstrichOR11	Functional	2	EU594798
Ostrich	OstrichOR12	Functional	1	EU594799
Ostrich	OstrichOR13	Functional	3	EU594800
Ostrich	OstrichOR14	Functional	1	EU594801
Ostrich	OstrichOR15	Functional	1	EU594802
Ostrich	OstrichOR16	Pseudogene	1	EU594803
Ostrich	OstrichOR17	Functional	1	EU594804
Ostrich	OstrichOR18	Functional	1	EU594805
Ostrich	OstrichOR19	Functional	3	EU594806
Ostrich	OstrichOR20	Functional	3	EU594808
Ostrich	OstrichOR21	Functional	2	EU594809
Ostrich	OstrichOR22	Pseudogene	1	EU594810
Ostrich	OstrichOR23	Functional	1	EU594811
Ostrich	OstrichOR24	Functional	1	EU594812
Ostrich	OstrichOR25	Functional	4	EU594813
Rhea	RheaOR1	Functional	1	EU594821
Rhea	RheaOR2	Functional	3	EU594832
Rhea	RheaOR3	Functional	6	EU594842
Rhea	RheaOR4	Functional	5	EU594843
Rhea	RheaOR5	Pseudogene	2	EU594844
Rhea	RheaOR6	Functional	2	EU594845
Rhea	RheaOR7	Pseudogene	1	EU594846
Rhea	RheaOR8	Functional	1	EU594847
Rhea	RheaOR9	Functional	1	EU594848
Rhea	RheaOR10	Functional	1	EU594822
Rhea	RheaOR11	Functional	1	EU594823

Rhea	RheaOR12	Functional	2	EU594824
Rhea	RheaOR13	Functional	1	EU594825
Rhea	RheaOR14	Pseudogene	1	EU594826
Rhea	RheaOR15	Functional	1	EU594827
Rhea	RheaOR16	Functional	1	EU594828
Rhea	RheaOR17	Pseudogene	3	EU594829
Rhea	RheaOR18	Pseudogene	2	EU594830
Rhea	RheaOR19	Pseudogene	1	EU594831
Rhea	RheaOR20	Functional	1	EU594833
Rhea	RheaOR21	Pseudogene	4	EU594834
Rhea	RheaOR22	Functional	1	EU594835
Rhea	RheaOR23	Functional	1	EU594836
Rhea	RheaOR24	Functional	2	EU594837
Rhea	RheaOR25	Functional	1	EU594838
Rhea	RheaOR26	Functional	1	EU594839
Rhea	RheaOR27	Functional	1	EU594840
Rhea	RheaOR28	Functional	2	EU594841

(b)

Common name	OR gene	Potentially functional/Pseudogene	Copies	GenBank accession number
Kaka	KakaOR1	Functional	1	EU594706
Kaka	KakaOR2	Functional	1	EU594717
Kaka	KakaOR3	Functional	3	EU594721
Kaka	KakaOR4	Functional	1	EU594722
Kaka	KakaOR5	Functional	2	EU594723
Kaka	KakaOR6	Functional	2	EU594724
Kaka	KakaOR7	Functional	1	EU594725
Kaka	KakaOR8	Functional	2	EU594726
Kaka	KakaOR9	Functional	2	EU594727
Kaka	KakaOR10	Pseudogene	1	EU594707
Kaka	KakaOR11	Functional	1	EU594708
Kaka	KakaOR12	Functional	2	EU594709
Kaka	KakaOR13	Functional	11	EU594710
Kaka	KakaOR14	Functional	1	EU594711
Kaka	KakaOR15	Pseudogene	2	EU594712
Kaka	KakaOR16	Functional	1	EU594713
Kaka	KakaOR17	Pseudogene	1	EU594714
Kaka	KakaOR18	Functional	2	EU594715
Kaka	KakaOR19	Functional	3	EU594716
Kaka	KakaOR20	Pseudogene	5	EU594718
Kaka	KakaOR21	Pseudogene	1	EU594719
Kaka	KakaOR22	Functional	1	EU594720
Kakapo	KakapoOR1	Functional	1	EU594728
Kakapo	KakapoOR2	Functional	1	EU594739
Kakapo	KakapoOR3	Functional	1	EU594750
Kakapo	KakapoOR4	Functional	1	EU594760
Kakapo	KakapoOR5	Functional	1	EU594761
Kakapo	KakapoOR6	Functional	1	EU594762
Kakapo	KakapoOR7	Functional	3	EU594763
Kakapo	KakapoOR8	Functional	2	EU594764
Kakapo	KakapoOR9	Functional	5	EU594765
Kakapo	KakapoOR10	Functional	1	EU594729
Kakapo	KakapoOR11	Functional	1	EU594730
Kakapo	KakapoOR12	Functional	4	EU594731
Kakapo	KakapoOR13	Functional	1	EU594732
Kakapo	KakapoOR14	Functional	1	EU594733
Kakapo	KakapoOR15	Functional	1	EU594734
Kakapo	KakapoOR16	Pseudogene	1	EU594735

Kakapo	KakapoOR17	Functional	1	EU594736
Kakapo	KakapoOR18	Pseudogene	1	EU594737
Kakapo	KakapoOR19	Functional	1	EU594738
Kakapo	KakapoOR20	Pseudogene	1	EU594740
Kakapo	KakapoOR21	Pseudogene	1	EU594741
Kakapo	KakapoOR22	Pseudogene	1	EU594742
Kakapo	KakapoOR23	Functional	1	EU594743
Kakapo	KakapoOR24	Pseudogene	1	EU594744
Kakapo	KakapoOR25	Pseudogene	1	EU594745
Kakapo	KakapoOR26	Functional	1	EU594746
Kakapo	KakapoOR27	Pseudogene	1	EU594747
Kakapo	KakapoOR28	Pseudogene	1	EU594748
Kakapo	KakapoOR29	Functional	1	EU594749
Kakapo	KakapoOR30	Pseudogene	1	EU594751
Kakapo	KakapoOR31	Functional	1	EU594752
Kakapo	KakapoOR32	Functional	1	EU594753
Kakapo	KakapoOR33	Functional	1	EU594754
Kakapo	KakapoOR34	Functional	2	EU594755
Kakapo	KakapoOR35	Functional	2	EU594756
Kakapo	KakapoOR36	Pseudogene	1	EU594757
Kakapo	KakapoOR37	Functional	1	EU594758
Kakapo	KakapoOR38	Pseudogene	1	EU594759
Kea	KeaOR1	Functional	1	EU594766
Kea	KeaOR2	Functional	1	EU594777
Kea	KeaOR3	Pseudogene	1	EU594788
Kea	KeaOR4	Functional	2	EU594790
Kea	KeaOR4	Functional	4	EU594791
Kea	KeaOR6	Functional	1	EU594792
Kea	KeaOR7	Functional	1	EU594793
Kea	KeaOR8	Functional	5	EU594794
Kea	KeaOR9	Functional	1	EU594795
Kea	KeaOR10	Functional	4	EU594767
Kea	KeaOR11	Functional	1	EU594768
Kea	KeaOR12	Functional	1	EU594769
Kea	KeaOR13	Functional	1	EU594770
Kea	KeaOR14	Functional	1	EU594771
Kea	KeaOR15	Functional	1	EU594772
Kea	KeaOR16	Functional	1	EU594773
Kea	KeaOR17	Functional	1	EU594774
Kea	KeaOR18	Functional	1	EU594775
Kea	KeaOR19	Functional	2	EU594776
Kea	KeaOR20	Pseudogene	3	EU594778
Kea	KeaOR21	Functional	1	EU594779
Kea	KeaOR22	Functional	2	EU594780
Kea	KeaOR23	Functional	1	EU594781
Kea	KeaOR24	Functional	1	EU594782
Kea	KeaOR25	Functional	2	EU594783
Kea	KeaOR26	Functional	1	EU594784
Kea	KeaOR27	Functional	1	EU594785
Kea	KeaOR28	Functional	1	EU594786
Kea	KeaOR29	Functional	5	EU594787
Kea	KeaOR30	Functional	1	EU594789

Chapter 4

Evidence for adaptive evolution of olfactory receptor genes in nine bird species

Silke S. Steiger, Andrew E. Fidler, Jakob C. Mueller and Bart Kempenaers

ABSTRACT

It has been suggested that positive selection, i.e. selection favoring a change in the protein sequence, plays a role in the evolution of olfactory receptor (OR) gene repertoires in fish and mammals. Olfactory receptors are seven-transmembrane domain (TM) proteins, members of the G-protein coupled receptor (GPCR) superfamily in vertebrate genomes, and responsible for odorant binding and discrimination. OR gene repertoires in birds are surprisingly large and diverse, suggesting that birds have a keen olfactory sense. Therefore, avian ORs may have evolved rapidly. The aim of this study is to investigate signatures of positive selection in avian ORs. Using maximum-likelihood methods that estimate the $d_N \backslash d_S$ ratios and that account for the effects of recombination, we show here that there is evidence for positive selection in partial OR coding sequences of nine bird species that are likely to have different olfactory requirements: 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*), the red jungle fowl (*Gallus gallus*), and the snow petrel (*Pagodroma nivea*). Positively selected codon sites were predominantly located in TM regions, which - in other vertebrates - are involved in odorant binding. Interestingly, the number of positively selected sites for a particular species correlated positively with the estimated number of OR genes encoded in that species' genome and the species' relative olfactory bulb size, which are possible correlates of olfactory ability. Our data suggest that (i) avian OR genes have been subjected to adaptive evolution, (ii) the extent of adaptive evolution differs between bird species and (iii) is higher in species that have good olfactory abilities.

INTRODUCTION

The molecular basis of the sense of smell is mediated by olfactory receptors (ORs), discovered by Richard Axel and Linda Buck in 1991 (Buck and Axel, 1991). Olfactory receptors are seven-transmembrane domain (TM) proteins that are expressed on olfactory sensory neurons in the olfactory epithelium (Mombaerts, 1999b; Young and Trask, 2002; Mombaerts, 2004; Niimura and Nei, 2006). OR genes are small (~1 kb) and intronless (Buck and Axel, 1991) and comprise one of the largest multigene families in vertebrate genomes (range: 100 – 2130 ORs in the pufferfish, *Fugu rubripes*, and the cow, *Bos taurus*, respectively) (Niimura and Nei, 2005; Niimura and Nei, 2007).

Two major evolutionary processes that are not mutually exclusive have shaped the OR gene family: gene duplications, possibly mediated by unequal recombination, and gene conversions (Young and Trask, 2002). How can the diversity of the OR gene repertoire be increased after gene duplication? It has been suggested that positive selection (i.e. selection favoring changes in the protein sequence) contributes to a diverse repertoire of OR genes in fish (Ngai et al., 1993; Kondo et al., 2002; Alioto and Ngai, 2005; but see Sun et al., 1999) and mammals (Hughes and Hughes, 1993; Singer et al., 1996; Gilad et al., 2000; Nielsen et al., 2005; Moreno-Estrada et al., 2008; but see Gimelbrant et al., 2004; Zhang et al., 2004b). Notably, positive selection events may greatly influence the function of ORs, because even a single amino acid mutation in the binding site of an OR may alter the ligand specificity, thereby allowing the OR to recognize certain odorant molecules with higher or lower affinities (Katada et al., 2005).

It is still widely believed that olfactory cues are less important for birds than visual or auditory cues (for review, see Roper, 1999; Hagelin, 2006; Hagelin and Jones, 2007). However, the discovery that the OR gene family is surprisingly large in birds 'seems to run counter to the textbook view that birds have a poor sense of smell' (International Chicken Genome Sequencing Consortium, 2004) and indicates that birds rely more on olfactory cues than previously thought. In addition, recent behavioural studies showed that olfaction plays a much more important part in the behavioural ecology of birds (Bonadonna and Nevitt, 2004; Amo et al., 2008; Gwinner and Berger, 2008; Nevitt et al., 2008). Thus, we hypothesized that avian OR genes have evolved rapidly and will show signs of positive selection.

Vertebrate OR genes have been classified based on sequence similarity into two distinct groups: group-α (previously termed class I genes) and group-γ (previously termed class II), (Freitag et al., 1995; Niimura and Nei, 2005). Interestingly, previous studies showed that birds have a unique subset of OR genes (group-γ-c) (Niimura and Nei, 2005; Steiger et al., submitted). Group-γ-c OR genes are very similar in sequence and cluster together in a lineage-specific pattern in a phylogenetic tree (Steiger et al., submitted). This lineage-specific clustering pattern may be explained by recent gene recombination or gene conversion events (Steiger et al., submitted).

The aim of the present study was threefold. First, we tested whether gene recombination and/or gene conversion events played a role in the evolution of avian group-γ-c OR genes. Second, we used three different maximum likelihood methods to identify the key amino acid residues on which positive selection has acted. In addition, we tested whether there is evidence for alignment-wide selection (i.e. evidence of positive selection operating within recombining fragments of the alignment). Third, we determined the level

of interspecific variability in the number of positively selected sites and tested whether this correlated with two possible correlates of olfactory capability, the estimated OR gene repertoire size and the relative olfactory bulb size (Bang and Cobb, 1968; Niimura and Nei, 2006). We chose to investigate group-γ-c ORs of nine bird species from seven orders because (i) the possible correlates of olfactory capability vary widely among those species and (ii) the species occupy different ecological niches (e.g. habitat, diet, and activity pattern differ among species) and are therefore likely to have different and possibly even species-specific olfactory requirements, which may be reflected in their OR gene repertoires.

MATERIALS AND METHODS

Study species

We analyzed partial group-γ-c OR coding sequences of nine bird species from seven different orders: 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*), the red jungle fowl (*Gallus gallus*), and the snow petrel (*Pagodroma nivea*). The species studied occupy various ecological niches and are therefore likely to have different olfactory requirements (for more details, see Steiger et al., submitted). In addition, two possible correlates of olfactory capability, the olfactory bulb ratio (OBR) and the number of OR genes encoded in a species' genome, vary widely among species (Edinger, 1908; Niimura and Nei, 2006; Steiger et al., submitted). The olfactory bulb ratio (OBR) is the ratio of the greatest diameter of the olfactory bulb to the greatest diameter of the cerebral hemisphere in per cent and varies up to 12 fold among the avian species studied (Bang and Cobb, 1968). Similarly, the estimated OR gene repertoire sizes were estimated to vary up to 6-fold between species (range: 106 – 667 OR genes in the galah and the kakapo, respectively) (Steiger et al., submitted).

Partial OR coding sequences, alignment, and phylogenetic analysis

For each bird species except the galah, 20 group-γ-c partial OR coding sequences were obtained from Steiger et al. (submitted). For the galah, only five sequences could be included. In brief, partial OR coding sequences were obtained by using PCR with degenerate primers designed to anneal to evolutionarily conserved coding sequences corresponding to the end of TM3 (forward primers) and TM7 (reverse primers) regions of the OR proteins (Steiger et al., submitted). Partial OR coding sequences were ~ 0.5 kb in size, which represents approximately half of the expected full OR coding sequence. The accession numbers of all sequences used in this study are given in supplementary Table 4.1. A phylogenetic tree that confirmed their classification as group-γ-c OR coding sequences is shown in supplementary Figure 4.1. For selection analysis, we only used sequences from potentially functional OR genes, i.e. sequences that did not contain a stop codon. Because only few partial OR coding genes from other gene families (group-γ, group-α) and hardly any pseudogenes were identified previously (Steiger et al., submitted), we restricted our analysis to potentially functional group-γ-c OR genes.

All partial coding OR sequences were codon-aligned using the default settings of ClustalW (Larkin et al., 2007), which is included in the MEGA 4 package (http://www.megasoftware.net/; Tamura et al., 2007), and visually inspected. The

alignment was 468 bp long and can be obtained from the authors on request. Partial coding OR sequences were then analysed separately for each species (see below).

Recombination analysis

We used the Genetic Algorithm for Recombination Detection (GARD) tool implemented in the Datamonkey web-interface (http://www.datamonkey.org/GARD/) (Pond and Frost, 2005a) with default settings to detect (i) the presence or absence of recombination and (ii) the location of recombination break points. Models of nucleotide substitution belonging to the general time reversible (GTR) family were chosen based on the model selection procedure available. GARD screened the alignment, located non-recombinant fragments and built a phylogenetic Neighbour-Joining (NJ) tree for each fragment (Kosakovsky Pond et al., 2006). Break points were identified with score improvements in information-based criteria, such as the small sample-corrected Akaike Information Criterion (c-AIC) (Kosakovsky Pond et al., 2006), and mapped onto a sequence logo derived from the alignment of all partial OR coding sequences used in this study (see below). A NEXUS file with partition information output created by GARD was subsequently used as input for the positive selection analysis (see below).

Positive selection analysis

Selection at the protein level is generally indicated by the rate of nonsynonymous and synonymous substitutions ($\omega=d_N\backslash d_S$) (Yang and Bielawski, 2000). Positive selection, neutral evolution and purifying selection (selection against deleterious alleles and thus, against changes in the protein sequence), is indicated by $\omega>1$, $\omega=1$ and $\omega<1$, respectively (Yang and Bielawski, 2000).

Previous studies have used phylogeny-based detection methods to estimate past selection in OR genes (Zhang et al., 2004b; Nielsen et al., 2005). Such phylogeny-based selection detection programs (e.g. PAML, Yang, 1997) generally base their analysis on a single phylogenetic tree. However, no single phylogenetic tree can adequately describe the evolutionary history of the whole sequence in the presence of recombination (Anisimova et al., 2003). As a consequence, these programs tend to overestimate the extent of positive selection among members of multi-gene families (for review, see Anisimova and Liberles, 2007). New methods have recently become available that allow to test for positive selection in the presence of recombination (Kosakovsky Pond et al., 2006), such as the single likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), random-effects likelihood (REL) and PARRIS [A (PAR)titioning approach for (R)obust Inference of (S)election] methods implemented in the Datamonkey web-interface (Pond and Frost, 2005b; Scheffler et al., 2006). In these methods, several phylogenies - one for each non-recombinant fragment in the alignment - can be used as input and thus, the evolution of non-recombinant fragments can be modelled more precisely, and a more accurate estimation of positive selection can be performed.

(a) Site-specific selection

To identify past selection on individual codons, we used SLAC, FEL and REL methods with default settings, implemented in the Datamonkey web-interface (Pond and Frost, 2005a). In brief, SLAC first fits a nucleotide model to the data and calculates a global ω ratio. Then, the ancestral sequences at each codon are reconstructed using maximum

likelihood. Finally, the expected and observed numbers of synonymous and non-synonymous substitutions are calculated to infer selection at each codon site. Significance is assessed by using a p-value derived from a two-tailed binomial distribution. As SLAC is a conservative test, the default setting with a significance level of 0.1 was used following Pond and Frost (2005b).

Whereas SLAC calculates the expected and observed numbers of synonymous and non-synonymous substitutions to infer selection, FEL directly estimates d_N and d_S based on a codon-substitution model. A likelihood-ratio test is used to assess significance. For FEL, the default setting with a significance level of 0.1 was used.

REL is an extension of the site-by-site positive selection analyses implemented in PAML (Yang, 1997). Notably, it allows the synonymous and nonsynonymous substitution rates to vary among codon sites (Pond and Frost, 2005b). It makes use of the Bayes Factors to determine a site as selected. For REL, a Bayes Factor >50 was applied (default conditions). In general, REL is more powerful than SLAC and FEL but it tends to have the highest rate of false positives (Pond and Frost, 2005b).

Finally, we applied the 'Integrative Selection Analysis' to determine the total number of positively selected codons, which were detected by at least one of the three methods (Pond and Frost, 2005b).

(b) Alignment-wide selection

To test for alignment-wide signatures of selection, i.e. evidence of positive selection operating within recombining fragments of the alignment, we used the program PARRIS, implemented in the Datamonkey web-interface (Scheffler et al., 2006). PARRIS allows tree topologies and branch lengths to change across detected recombination break points. In addition, the synonymous substitution rates are allowed to vary across codon sites (Scheffler et al., 2006). First, a null model (no selection) was fitted to the data, followed by an alternative (selection) model. A likelihood ratio test can then be used to compare models and to test whether there is evidence for positive selection (Scheffler et al., 2006).

Sequence logo generation

The WebLogo program (http://weblogo.berkeley.edu/logo.cgi) was used to generate a sequence logo from the amino acid alignment of all partial OR coding sequences. The height of each symbol corresponds to the relative frequency of the amino acid at a particular position. The approximate positions of TMs, extracellular domains (EC) and intracellular domains (IC) were located based on a combination of the results of a TM prediction online tool (BCM Search Launcher; http://searchlauncher.bcm.tmc.edu/seq-search/struc-predict.html) and the suggested location of mammalian TMs (Liu et al., 2003). Note that the exact number and precise placement of the TMs has not been experimentally verified and should thus be treated with caution.

Correlation of positively selected sites with predictors of olfactory ability

We tested whether the total number of positively selected codons correlated with the estimated number of OR genes encoded in the genome and the OBR of that species, which can be considered indicators of olfactory ability (data from Steiger et al.,

submitted; Bang and Cobb, 1968; Hagelin, 2004). To control for phylogenetic non-independence, we calculated phylogenetically independent contrasts (PIC; Felsenstein 1985) using the PDAP:PDTree module of Mesquite (Midford et al., 2005; Maddison and Maddison, 2006). For the topology of the tree, see Steiger et al., submitted. Note that for the black coucal, the OR gene repertoire size has not been estimated (Steiger et al., submitted).

RESULTS

Recombination analysis

We found evidence for recombination in partial OR coding sequences in all bird species studied, with between 2 and 4 recombination break points (Table 4.1). Note that the locations of the recombination break points were very similar among bird species; in six cases, they were identical for at least two species (Table 4.1). Recombination break points were located predominantly in EC2, in TM5, and TM6 (Figure 4.1).

Positive selection analysis

In general, global d_N/d_S ratios were low, indicating that avian OR genes are under purifying selection (Table 4.2). SLAC detected no positively selected sites in three species, and between one and three sites in the remaining 6 species (Table 4.2). FEL and REL analysis identified between 0 and eight, and between 0 and 15 codon sites under positive selection, respectively (Table 4.2). In both the SLAC and REL analysis, the brown kiwi had the largest number of positively selected sites. In the FEL analysis, the largest number of positively selected sites could be detected in the kakapo and the mallard. All sites that showed a signature of positive selection were mapped onto a sequence logo derived from the alignment of all OR genes used in this study (Figure 4.1). Six out of seven identified amino acids that were identified with all three methods cluster in the highly variable TM regions, i.e. TM4, TM5 and TM6 (Figure 4.1). Similarly, seven out of eleven amino acids that were identified in more than one species were located in TM regions. The PARRIS test revealed evidence for positive selection in the alignment in the brown kiwi, but not in any of the other species (Table 4.3). However, note that the p-values were low (0.18-0.25) for the kakapo and the red jungle fowl.

Table 4.1
Evidence for recombination in avian partial OR coding sequences. The change in the Akaike Information Criterion (**Δc-AIC**) and the number and the location of recombination break points on the nucleic and on the amino acid level estimated by GARD. **Δc-AIC** indicates the difference in the AIC between the non-recombination model (single tree model) and the best recombination model. Bold break point locations indicate break points that were observed in at least two species.

Common name	Δc-AIC	Number of break points	Break point location[1]
Black Coucal	751.33	3	**147 (49)**, **241 (81)**, 351 (117)
Blue tit	534.6	3	**144 (48)**, 233 (78), 363 (121)
Brown kiwi	208.47	3	124 (42), 248 (83), **346 (116)**
Canary	392.45	3	**147 (49)**, **234 (78),** 352 (118)
Red jungle fowl	541.38	4	87 (29), 168 (56), **241 (81)**, **346 (116)**
Galah	78.36	3	199 (67), **234 (78)**, 307 (103)
Kakapo	616.79	3	**144 (48)**, 249 (83), 373 (125)
Mallard	279.55	2	**144 (48)**, **374 (125)**
Snow petrel	420.05	3	129 (43), 243 (81), **374 (125)**

[1]Numbers indicate break point locations on the nucleic acid level. Numbers in parentheses show corresponding amino acid positions.

Figure 4.1 (opposite page)
Sequence logo of all avian partial coding group-γ-c OR genes. The Y and X axis indicate information content and amino acid position, respectively. Symbol height indicates the relative frequency of each amino acid at that position. Transmembrane domains (TM), intracellular (IC) and extracellular (EC) domains are indicated. Note that the exact number and precise placement of the TMs has not been experimentally verified and is thus speculative. Asterisks above amino acids indicate the location of recombination break points. Triangles, circles and squares under amino acids indicate positively selected sites determined by SLAC, FEL, and REL, respectively. Open and closed symbols indicate sites that were observed in one and at least two bird species, respectively.
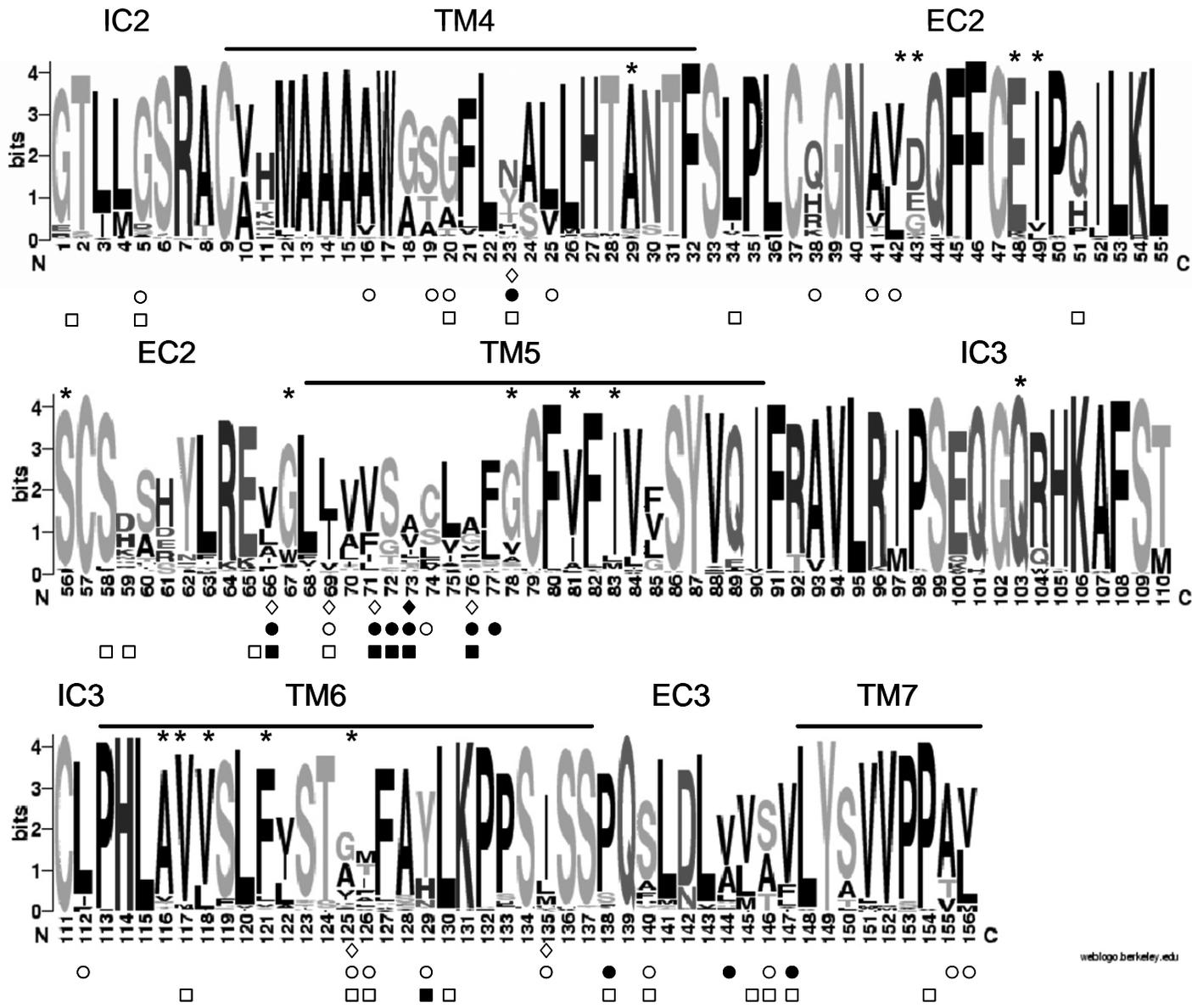
Table 4.2 (opposite page)
Number and location of codons under positive selection in avian partial OR coding sequences that belong to the group-γ-c OR gene clade. The single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), and random effects likelihood (REL) methods were used to identify positively selected codons (see Methods for details). Underlined and bold codon sites indicate codons that were identified with two and three methods, respectively.

Correlation of positively selected sites with predictors of olfactory ability

Across species, the total number of positively selected codons identified with the integrative selection analysis correlated positively with the estimated number of OR genes and with the OBR (Table 4.4), but this was only significant for the former. Similarly, the number of positively selected codons identified with either SLAC, FEL or REL were always positively correlated with the estimated number of OR genes and the OBR, although the correlations were not always significant (Table 4.4). Note that qualitatively similar results were obtained when (i) we did not control for phylogeny, (ii) a more recent phylogeny (Steiger et al., submitted) was applied, and (iii) the species with only few OR sequences (galah) was excluded from the analysis (details not shown).

DISCUSSION

In this study, we found evidence that gene recombination and/or gene conversion had an impact on the evolution of the avian OR gene repertoires and that positive Darwinian selection contributed to the molecular evolution of OR genes in birds.

Previous studies estimated past selection in OR genes using phylogeny-based methods, although these methods suffer from the problem that the extent of positive selection among members of multi-gene families may be overestimated in the presence of recombination (for review, see Anisimova and Liberles, 2007). It is important to explicitly take recombination/gene conversion into account before testing for positive selection because both gene conversion and recombination events have been detected in the OR gene family. For example, gene conversion events that encompass the hypervariable TM regions 3, 4 and 5 and the second extracellular loop (EC2, region between TM4 and TM5) have been shown in primate OR genes (Sharon et al., 1999). Recombination events were also detected at the TM4 region in the Japanese medaka fish, *Oryzias latipes* (Kondo et al., 2002). To our knowledge, our study is the first that controlled for such recombination/gene conversion events before the $d_N\backslash d_S$ ratio was calculated. Indeed, we detected gene recombination events in EC2, in TM5, and TM6 in all bird species investigated (Figure 4.1) and thus, support previous studies suggesting that recombination events contributed to vertebrate OR gene evolution. On average, more positively selected sites were identified with each method when no control for recombination was carried out, supporting the notion that the extent of positive selection is overestimated in such cases. However, the difference was not significant (details not shown).

Common name	Global $d_N\backslash d_S$	SLAC		FEL		REL		Integrative Selection Analysis
		Number	Location	Number	Location	Number	Location	Total number
Black coucal	0.47	2	73, *135*	5	23, 72,73, 135, 147	2	72, 147	5
Blue tit	0.63	0		4	20, 77, 129, 155	6	20, 73, 76, 125, 126, 129	8
Brown kiwi	0.65	3	**69**, **73**, **76**	5	23, **69**, 71, **73**,**76**	15	1, 5, 23, 34, 51, 66, **69**, 71, 72, **73**, **76**, 129, 140, 145, 146	15
Canary	0.55	0		4	23, 41, 42, 147	0		4
Galah	0.60	0		0		1	59	1
Kakapo	0.57	1	23	8	19, 23, 38, 66, 71, 72, 138,146	1	66	8
Mallard	0.44	1	**71**	8	5, 16, 23, 66 , **71**, 74, 126, 144	1	**71**	8
Red jungle fowl	0.46	1	**66**	6	25, **66**, 76, 77, 138, 140	7	58, 65, **66**, 117, 130, 138, 154	11
Snow petrel	0.55	1	125	4	112, 125, 144, 156	0		4

Table 4.3
Evidence for alignment-wide positive selection in avian partial OR coding sequences. The PARRIS method was used to test for adaptive evolution (see methods for details). Log likelihoods (Log (L)) of the null model and the alternative model are shown. The Likelihood Ratio Tests (LRT) of the null model versus the alternative model was significant in the brown kiwi and indicates evidence for positive selection.

Common name	Null model Log (L)	Alternative model Log(L)	LRT	p-value	Evidence for positive selection
Black Coucal	-3376.11	-3376.11	0	1	no
Blue tit	-2224.91	-2224.85	0.12	0.94	no
Brown kiwi	-3234.09	-3226.56	15.06	<0.001	yes
Canary	-2505.69	-2505.69	0	1	no
Galah	-1158.85	-1158.72	0.26	0.89	no
Kakapo	-2599.83	-2598.13	3.4	0.18	no
Mallard	-3487.7	-3487.7	0	1	no
Red jungle fowl	-3312.92	-3311.52	2.8	0.25	no
Snow petrel	-2570.1	-2570.41	-0.62	1	no

Table 4.4
The relationship between the number of positively selected sites and the estimated number of OR genes and the olfactory bulb size ratio (OBR), respectively (see Methods for details). Shown are the Pearson's correlation coefficient and the corresponding P-value.

Method	Estimated number of OR genes (N = 8)		OBR (N=9)	
	r	p	r	p
SLAC	0.32	0.44	0.71	<0.05
FEL	0.87	<0.01	0.62	0.08
REL	0.32	0.43	0.17	0.65
Integrative Selection Analysis	0.84	<0.01	0.54	0.13

Several studies showed that there is evidence for positive selection in the OR gene family (e.g. Hughes and Hughes, 1993; Ngai et al., 1993; Singer et al., 1996; Sun et al., 1999; Kondo et al., 2002; Zhang et al., 2004b; Alioto and Ngai, 2005). Nevertheless, it has been argued that a small degree of positive selection may not be important, because the ability of olfaction may be largely determined by the number of OR genes (Niimura and Nei, 2006). Alternatively, a small degree of positive selection may be adaptive when structurally important codons are under selection (Katada et al., 2005). Indeed, codons identified as being influenced by positive selection were frequently located in those domains that are directly related to the function of the gene (i.e. the odorant binding sites). Binding sites of odorants are located in TM3, TM4, TM5 and TM6 in mouse, *Mus musculus*, (Katada et al., 2005) and rat, *Rattus norvegicus*, ORs (Hall et al., 2004). Previous studies have documented positive selection in TM3 and TM4 in the channel catfish, *Ictalurus punctatus*, (Ngai et al., 1993) and in the zebrafish, *Danio rerio* (Alioto and Ngai, 2005). Similarly, we found that many positively selected sites were located in TM domains. For example, seven out of eight codons identified with the most conservative SLAC method, were located in TM domains. Four out of the seven codons were located in TM5, a domain which forms the putative ligand-binding pocket (Floriano et al., 2004). Furthermore, two of these codons showed signatures of positive selection in at least two bird species, independently of the method. Note that the only positively selected codon identified with SLAC that was not in a TM domain is located very close to TM5. Because the location of the TM domains has not been experimentally verified and the 'true' position may thus slightly change, it is possible that this codon is also located within TM5. The results from the FEL and REL methods did not yield such a clear pattern, i.e. the location of positively selected sites was not restricted to the TM domains. However, the sites outside TM regions were generally detected with (i) only one method and (ii) in only one species, suggesting that these results may not be as reliable as the SLAC results. Taken together, our results suggest that the sites that show signatures of positive selection are functionally relevant. Structure-function relationships of avian OR genes have not been investigated and might be worthy of future exploration.

We showed that positive selection is more prevalent in species with a larger estimated OR gene repertoire and with a larger OBR (Table 4.4). Strikingly, the brown kiwi, a nocturnal and flightless bird that is in known to highly rely on olfactory cues (Wenzel, 1968; Bang, 1971) showed the highest number of positively selected sites. In addition, we found evidence for alignment-wide positive selection in the brown kiwi. This is generally difficult to detect because sites that are under positive selection are masked if the majority of sites are under purifying selection (Yang and Bielawski, 2000). Thus, this suggests that in the brown kiwi many sites are under positive selection. In contrast, we found generally fewer positively selected sites in the canary, the galah and the blue tit. It is currently highly debated whether olfactory cues are of major importance for those species (particularly for species of the order *Passeriformes*) (Amo et al., 2008). In any case, our results coincide with the anatomical facts that the relative sizes of the olfactory bulb and the estimated numbers of OR genes are reduced in those three species (Bang and Cobb, 1968). In conclusion, our results suggest that the strength of directional selection operating on intact OR genes differs among species.

We showed that positive selection may have contributed to the fast diversification of avian group-γ-c ORs. It would be interesting to investigate whether other OR gene

families (e.g. group-α or group-γ OR genes) show similar signs of recombination and positive selection. In addition, future studies could investigate whether pseudogenes evolve neutrally, as would be expected and as has been shown for e.g. mammalian vomeronasal receptors, which are the major receptors for pheromones in vertebrates (Mundy and Cook, 2003).

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Figure 4.1

Unrooted Neighbour-Joining (NJ) phylogenetic tree derived from an alignment of avian partial OR coding sequences with red jungle fowl (*Gallus gallus*) OR sequences that were obtained from Niimura and Nei (2005) (sequences listed in Supporting data set 8, http://www.pnas.org/cgi/content/full/0501922102/DC1/13, n = 78). The NJ method was used to generate the NJ tree from Poisson correction distances using the MEGA software (http://www.megasoftware.net/). Bird species included in the NJ phylogenetic tree are the black coucal (*Centropus grillii*, 20 sequences, ●), the blue tit (*Cyanistes caeruleus*, 20 sequences, ●), the brown kiwi (*Apteryx australis*, 20 sequences, ●), the canary (*Serinus canaria*, 20 sequences, ●) the galah, (*Eolophus roseicapillus*, 5 sequences, ●), the kakapo (*Strigops habroptilus*, 20 sequences, ●), the mallard (*Anas platyrhynchos*, 20 sequences, ●), the red jungle fowl (*Gallus gallus*, 20 and 78 sequences, ● and ●) and the snow petrel (*Pagodroma nivea*, 20 sequences, ●). Red jungle fowl sequences that were obtained from Niimura and Nei (2005) are indicated with black dots, while red jungle fowl sequences amplified using the PCR based approach are indicated with yellow dots. The large γ-c OR clade is shaded in grey. The scale-bar indicates the number of amino acid substitutions per site. The NJ phylogenetic tree confirmed that all partial avian OR sequences used for the positive selection analysis are members of the group-γ-c OR genes.
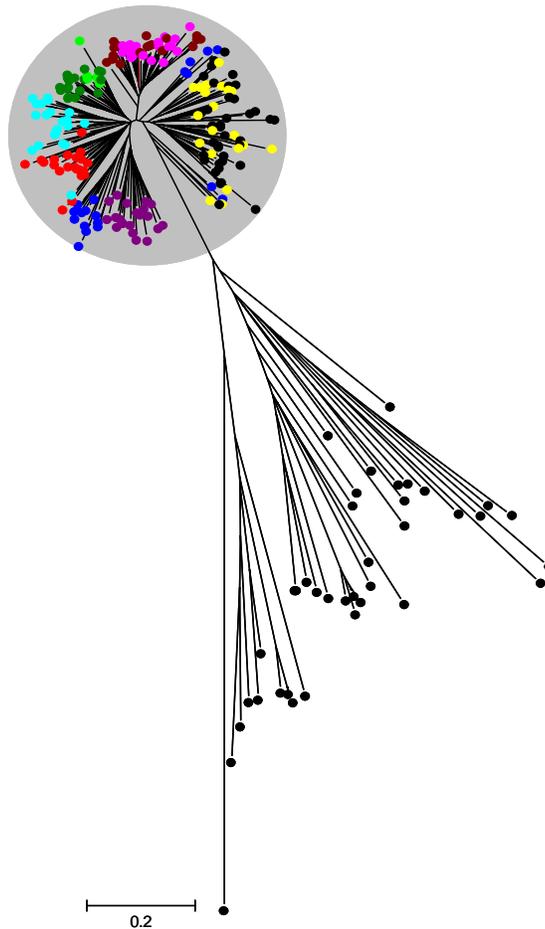
Table 4.1
Overview of the bird species and the partial group-γ-c OR coding sequences investigated in this study.

Common name	Latin name	Order	Number of partial group-γ-c OR coding sequences	Accession numbers
Black coucal	*Centropus grillii*	Cuculiformes	20	EF426980, EF426989, EF426990, EF426991, EF426992, EF426994, EF426995, EF426997,EF426998, EF426981, EF426999, EF42700, EF427001, EF427002, EF427003, EF426983, EF426984, EF426986, EF426987, EF426988
Blue tit	*Cyanistes caeruleus*	Passeriformes	20	EF426863, EF426873, EF426874, EF426875, EF426876, EF426877, EF426878, EF426879,EF426880, EF426881, EF426864, EF426882, EF426883, EF426884, EF426865, EF426866, EF426867, EF426868, EF426869, EF426871
Brown kiwi	*Apteryx australis*	Apterygiformes	20	EF427169, EF427178, EF427179, EF427180, EF427182, EF427185, EF427187, EF427170, EF427188, EF427190, EF427192, EF427194, EF427196, EF427197, EF427199, EF427172, EF427173, EF427174, EF427176, EF427177
Canary	*Serinus canaria*	Passeriformes	20	EF426925, EF426934, EF426935, EF426936, EF426938, EF426939, EF426940, EF426942, EF426943, EF426944, EF426945, EF426946, EF426947, EF426949, EF426950, EF426951, EF426952, EF426955, EF426929, EF426932
Galah	*Eolophus roseicapillus*	Psittaciformes	5	EF427041, EF427045, EF427046, EF427048, EF427049
Kakapo	*Strigops habroptilus*	Psittaciformes	20	EF427122, EF427125, EF427126, EF427127, EF427128, EF427130, EF427131, EF427114, EF427133, EF427134, EF427135, EF427136, EF427137, EF427139, EF427140, EF427115, EF427117, EF427118, EF427119, EF427120
Mallard	*Anas platyrhynchos*	Anseriformes	20	EF427237, EF427246, EF427247, EF427248, EF427249, EF427250, EF427251, EF427252, EF427253, EF427254, EF427255, EF427256, EF427259, EF427260, EF427261, EF427240, EF427242, EF427243, EF427244, EF427245
Red Jungle fowl	*Gallus gallus*	Galliformes	20	EF427067, EF427077, EF427078, EF427079, EF427080, EF427081, EF427082, EF427083, EF427084, EF427085, EF427068, EF427086, EF427087, EF427069, EF427070, EF427071, EF427072, EF427073, EF427074, EF427075
Snow petrel	*Pagodroma nivea*	Procellariiformes	20	EF427298, EF427308, EF427309, EF427310, EF427311, EF427312, EF427313, EF427314, EF427315, EF427316, EF427317, EF427320, EF427321, EF427322, EF427301, EF427302, EF427303, EF427304, EF427305, EF427306

Chapter 5

Detection of olfactory receptor transcripts in bird testes

Silke S. Steiger, Andrew E. Fidler and Bart Kempenaers

ABSTRACT

The sense of smell is mediated through olfactory receptors (ORs) expressed in olfactory sensory neurons of the olfactory epithelium. Interestingly, some OR genes also function in another context: they are expressed in non-olfactory testicular tissue and in sperm of mammals and fish where they mediate sperm flagellar motility. The presence of OR transcripts in testicular tissue of both mammals and fish suggests that this is a conserved trait among vertebrates. In birds, sperm competition is widespread and its outcome depends, in part, on sperm motility. Thus, avian testicular OR gene expression might be particularly interesting to study, especially in the context of current ideas on post-copulatory sexual selection. Using reverse transcription (RT)-PCR with degenerate primers specific for OR genes, and subsequent cloning, we here demonstrate that multiple OR gene transcripts are present in chicken (*Gallus gallus domesticus*) testes. Moreover, we show that they belong to the class-γ OR gene clade. We discuss the potential significance and evolutionary implications of avian testicular OR gene expression.

INTRODUCTION

Olfactory receptors (ORs) are involved in the detection and discrimination of odours in both invertebrates and vertebrates (Buck and Axel, 1991). In mammals, OR genes constitute the largest gene family in the genome (for review, see Niimura and Nei, 2006).

OR gene expression is not restricted to the olfactory tissue. Interestingly, transcripts of a subset of OR genes have been detected in testis and sperm (e.g. Parmentier et al., 1992) and other non-olfaction related tissues of both mammals and fish (for review, see Young and Trask, 2002), suggesting that the functions of olfactory receptors are not restricted to the context in which they were first characterized, namely olfaction.

Although the functions of OR genes in many non-olfactory tissues remain unclear, mammalian testicular/sperm OR proteins have been localized to the sperm flagellum midpiece and have been shown to mediate flagellar motility (Spehr et al., 2003). Hence, sperm expressed OR genes may play an important role in sperm-egg chemotaxis/communication (Spehr et al., 2004b).

While OR genes have been intensively studied in a wide range of species, comparatively little is currently known about avian OR genes. In the red jungle fowl (*Gallus gallus*) genome, recent estimates of the OR gene repertoire size range between 229 to 558 paralogues (International Chicken Genome Sequencing Consortium, 2004; Niimura and Nei, 2005; Lagerstrom et al., 2006), suggesting that birds may rely more on their sense of smell than is generally thought. Vertebrate OR genes have been classified into two distinct groups: class-α (previously termed class I) and class-γ (previously termed class II) (Freitag et al., 1995; Niimura and Nei, 2005). A subset of the class-γ OR genes, termed class-γ-c, is greatly expanded in avian genomes (Lagerstrom et al., 2006; Niimura and Nei, 2005; Steiger et al., submitted).

Notwithstanding accumulating evidence to the contrary, birds are still widely believed to lack a well-developed sense of smell (for reviews, see Roper, 1999; Hagelin, 2006). Thus, it is perhaps not surprising that birds have been a neglected group with respect to the study of OR gene expression. In this study, we investigated whether OR gene transcripts could be detected in avian (chicken, *Gallus gallus domesticus*) testes. We demonstrate that multiple OR transcripts can be detected in chicken testes RNA and show that they fall into the γ-OR clade.

MATERIALS AND METHODS

Isolation of RNA

Samples of chicken testis, liver and small intestine tissue were collected after sacrifice and immediately frozen in liquid nitrogen before storage at -80ºC. We homogenized tissue samples using ceramic beads (Precellys® Kit CK 14, Bertin Technologies, France) and used a commercial kit to isolate total RNA (RNeasy, Qiagen, Hilden, Germany). Total RNA was digested with RNAse-free DNAse (DNaseI, Qiagen).

Reverse transcription (RT)-PCR

We used two different pairs of degenerate primers, targeting either (a) the non-γ-c OR clade sequences (forward primer ORFor1/ reverse primer ORRev1) or (b) the large γ-c OR clade (ORFor2/ORRev2). Primer pairs were designed to anneal to evolutionarily conserved transmembrane (TM) 3 and TM7 coding regions. For more details, see supplementary Methods. Both primer pairs were predicted to generate products of ~ 0.5 kb representing approximately half the expected OR full-length coding sequence. To control for cDNA quality, we amplified a chicken β-actin partial coding sequence.

Sequencing and sequence analyses

Amplified products were excised from agarose gels and purified before ligation into a T-tailed cloning vector (pGemT-easy, Promega, Madison, U.S.A.). Plasmids were purified from individual transformed DH5α colonies by alkaline lysis and sequenced. We deleted vector and primer sequences from the raw sequences and those sequences sharing ≥98.5% identity were considered to be replicated amplifications from the same OR gene transcript and therefore were combined to form a consensus sequence. Sequence identities were calculated using the 'sequence identity matrix' function of BioEdit (Hall, 1999; http://www.mbio.ncsu.edu/BioEdit/bioedit.html). BLAT searches of the red jungle fowl genome (Build 2.1, released in May 2006) with default parameter settings were performed in an attempt to identify homologous regions in the red jungle fowl genome using the UCSC Genome Browser (http://genome.ucsc.edu/).

Phylogenetic tree construction

We aligned amino acid sequences predicted from the chicken PCR products with the TM3–TM7 predicted protein sequences from 78 previously reported, putatively functional, red jungle fowl OR receptor genes (Niimura and Nei, 2005) using ClustalX (Thompson et al. 1997) with default settings. We used the Neighbour-Joining (NJ) method and Poisson-distances to construct phylogenetic trees using the MEGA software package (version 3.0; http://www.megasoftware.net/). The reliability of the phylogenetic tree was evaluated with 1000 bootstrap repeats.

RESULTS

Using a PCR primer pair targeting non-γ-c OR genes (ORFor1 / ORRev1), we amplified a discrete band of ~0.5 kb from chicken testis cDNA (Figure 5.1a). To control for the possibility that the products were amplified from genomic DNA potentially contaminating the RNA preparations, we carried out two controls: (i) reactions in which the template was total RNA, without a reverse transcriptase step (data not shown) and (ii) reactions in which the template was cDNA derived from chicken liver and small intestine total RNA (Figure 5.1a). In neither case was a PCR product generated. In contrast, the PCR primer pair targeting the γ-c OR clade (ORFor2 / ORRev2) did not generate amplification products from any of the chicken cDNA templates including testis (Figure 5.1a). It is possible that this result may simply indicate that the reaction conditions were not suitable for the ORFor2 / ORRev2 primers. However, this seems highly unlikely because the ORFor2 / ORRev2 primers amplified a product of the expected size from a chicken genomic DNA template

(Figure 5.1b). The β-actin primers amplified a product of the expected size (~ 0.15 kb) from all three tissue chicken cDNAs (Figure 5.1c).

Plasmids carrying the ORFor1 / ORRev1 PCR amplification products from testes cDNA were isolated from 20 independent transformed colonies and sequenced. We identified six distinct partial OR coding sequences, here denoted GgTestesORA-F (GenBank accession number **EU583984-EU583989**). Sequences GgTestesOR-B, -D and -E are consensus sequences derived from 12, 3 and 2 plasmids respectively, while sequences GgTestesOR-A, -C and -F were obtained from a single plasmid. We detected no premature stop codons within the predicted open reading frames in any of the OR partial coding sequences, suggesting that all six sequences were derived from transcripts encoding functional OR proteins. The six OR sequences were between 50-97 % and 36-95% identical on the DNA and predicted protein levels, respectively (supplementary Table 5.1).

Phylogenetic analysis placed all six OR sequences within the evolutionarily older, non-γ-c OR clade (Figure 5.2). Note that three OR sequences (GgTestesOR-A, -B and -E) cluster very closely in the phylogenetic tree. Sequence identities among these three sequences range from 93-97% and 88-95% on the DNA and predicted protein levels, respectively (supplementary Table 5.1). Furthermore, all three sequences align with the same part of the red jungle fowl genome (Build 2.1) as the best 'hit' when used as query sequences in BLAT searches, although the alignments themselves have different percent identities and differ when inspected by eye (supplementary Table 5.2). Surprisingly, every difference in a GgTestesOR-A versus GgTestesOR-B comparison is a C versus T. We examined the electropherogram raw data and see no reason to doubt the quality of the sequence data. In addition, such high rates of C/T differences were not observed in any other pairwise comparisons of sequences derived from the same PCR reaction indicating that they are not an artifact of the experimental procedure.

(a)

(b)

(c)
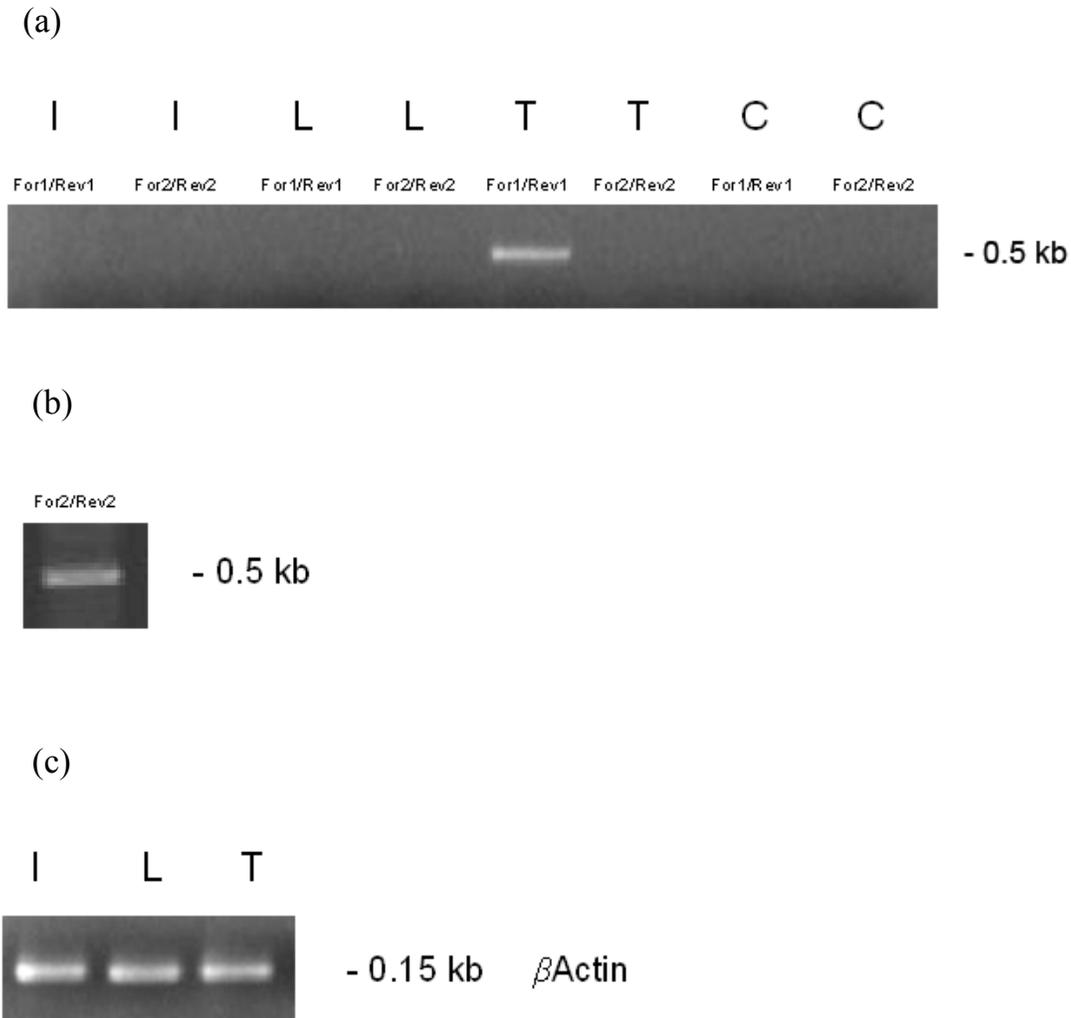
Figure 5.1
PCR amplification products generated by RT-PCR analysis of chicken (*Gallus gallus domesticus*) tissue (small intestine, liver, testis) cDNAs. (a) OR transcript detection: ORFor1/ORRev1 pair directed at non-γ-c group OR genes, ORFor2/ORRev2 pair directed at group-γ-c OR genes. (b) OR gene detection in chicken genomic DNA using the ORFor2/ORRev2 primer pair directed at group-γ-c OR genes (c) β-actin transcript detection Abbreviations: T, testis; I, small intestine; L, liver; C, negative control. Estimated sizes (kb) of the products are indicated.

Figure 5.2 (opposite page)
Neighbour-Joining (NJ) phylogenetic tree generated from an alignment of the predicted proteins derived from 78 putatively functional red jungle fowl (*Gallus gallus*) OR genes (sequences from Niimura and Nei, 2005) and the six testes-expressed chicken (*Gallus gallus domesticus*) OR genes generated in the work (black dots). The scale-bar indicates the number of amino acid substitutions per site. Numbers on tree branches show bootstrap values obtained from 1000 replications (only values >90% are shown).

92
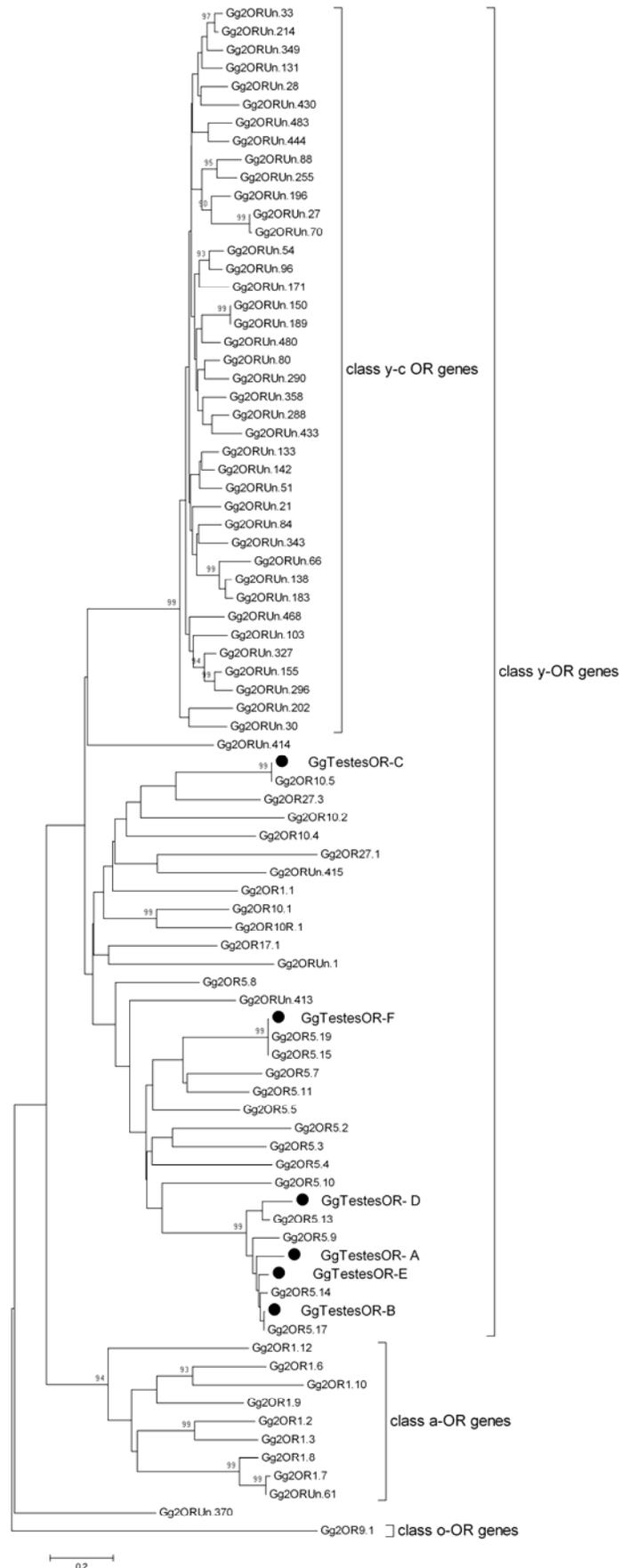
DISCUSSION

Using RT-PCR and subsequent cloning / sequencing, we here show that a minimum of six distinct OR genes are transcribed in the chicken testis. This number is probably an underestimate because we sequenced only 20 plasmids. Further RT-PCR / sequencing work, or perhaps microarray analyses, could be used to determine whether chicken and other bird species express as many different OR transcripts in their testes as do mammals (e.g., 66 different OR transcripts have been described in mice testis RNA; Zhang et al., 2004a).

Although the testis RNA used in this work was derived from a single bird, we cannot exclude the possibility that one pair of the three highly homologous GgTestesOR-A, -B and -E sequences are allelic. However, this seems unlikely for the following reasons: firstly, the percent identity values of the A, B and E sequences are all below the, *albeit* arbitrary, 'cut-off' values used in previous studies to designate distinct OR genes (e.g. Nef et al., 1996). Secondly, the frequencies of nucleotide mismatches among those three genes (i.e. between 13 and 34 mismatches in approximately half of the coding region) are much higher than those previously reported for mammalian OR gene alleles (i.e. maximally 11 mismatches in full-length coding regions) (Tacher et al., 2005; Moreno-Estrada et al., 2008). Thirdly, the current red jungle fowl genomic sequence (Build 2.1) is still a draft, *albeit* an advanced one, so OR sequences may be incorrect and / or absent. Indeed, the GgTestesOR-A, -B and -E sequences all align, as their best 'hit', with the same region of red jungle fowl genome (Build 2.1) but this region of chromosome 5 contains approximately 2500 sequence gaps (i.e. 4.6% of chromosome 5, mean length approx 1.0 kb) and the largest gap is estimated to be approximately 1.5 Mb (UCSC Genome Browser; http://genome.ucsc.edu/). As homologous mammalian OR genes tend to occur in clusters (Glusman et al., 2001; Godfrey et al., 2004), it is possible that additional OR sequences will be found in the chromome 5 'gaps' in further drafts of the red jungle fowl genome.

Further, it is becoming apparent that the genome of the domesticated chicken (*Gallus gallus domesticus*) is probably not a simple, direct derivative from that of the red jungle fowl (*Gallus gallus*) but, rather, may have been formed by hybridization of *Gallus gallus* and one or more other members of the genus *Gallus* (Eriksson et al. 2008). If this hybridization scenario is correct then genuinely orthologous OR sequences from the chicken (*Gallus gallus domesticus*) and *Gallus gallus* may not be identical. In addition, concerted evolution may cause OR genes to become very similar presenting a further challenge to the BLAT search program (Sharon et al., 1999, Steiger et al., submitted).

We further demonstrated that the six OR transcripts amplified can be placed in the OR group-γ, but not in a subclade, group-γ-c, whose expansion is apparently specific to bird genomes (Niimura and Nei, 2005, Steiger et al., submitted). In mammals, testes-expressed OR genes do not appear to belong to a single clade (Vanderhaeghen et al., 1997). Similarly, we have found that the six chicken testes-expressed ORs identified in this study were distributed throughout the group-γ clade. Nonetheless there is some evidence of a bias in the group-γ OR genes expressed in the chicken testes, because four sequences (GgTestesOR-A, -B, -D and -E) are placed in a single subclade. Furthermore, three of these sequences (GgTestesOR-B, -D and -E) were recovered multiple times within the data-set, which may indicate higher levels of expression

compared with the other three OR genes which were found in a single plasmid. This observation should not be over-interpreted as biases inherent in the PCR make extrapolations between ratios in the final PCR products and those in the original template problematic. Since we did not detect any testes-expressed OR genes from the γ-c clade we speculate that this clade may be predominantly, or exclusively, used in olfaction, whereas ORs in other clades may have additional, non-olfactory functions. If correct, one would expect differing evolutionary selective pressures to be acting on these two hypothetically functionally different categories. OR genes only used in olfaction are expected to be under comparatively relaxed, or perhaps positive, selection and therefore may evolve more rapidly than OR genes that are constrained by having multiple functions. Indeed, there is evidence from mammals that testes-expressed ORs are evolutionarily more conserved than ORs expressed in olfactory tissue (Branscomb et al., 2000).

Spehr et al. (2003) showed that testicular/sperm OR proteins are localized in the sperm flagellum midpiece and influence flagellar motility. This finding suggests that sperm expressed OR genes may play an important role in sperm-egg chemotaxis/communication (Spehr et al., 2004a). It is reasonable to hypothesize that these findings can be extrapolated to birds. Interestingly, in birds sperm motility positively correlates with fertilization success and thus, male fitness (Birkhead et al., 1999). Sperm motility is of particular importance when ejaculates of different males compete for fertilization, a phenomenon described as sperm competition (Parker, 1970). In many avian species, females mate with more than one male (Petrie and Kempenaers, 1998), hence, avian sperm competition is common. However, little is currently known about the underlying mechanisms of post-copulatory selection. Thus, further studies of the evolution and function of OR genes expressed in avian testes may prove a fruitful area of research.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Reverse transcription (RT)-PCR

In this study two different pairs of degenerate primers were used, targeting either (a) the non-γ-c OR clade sequences (forward primer ORFor1: 5'- ATGGCITAYGAYMGITA – 3', Nef et al., 1996 / reverse primer ORRev1: 5'- TADATIAGIGGRTTIAGCATIGG -3', Steiger et al. submitted) or (b) the large γ-c OR clade (forward primer ORFor2: 5' – ATCTGYAARCCIYTICAYTA –3', Steiger et al. submitted / reverse primer ORRev2: 5'- ARISWRTADATRAAIGGRTT -3', Freitag et al., 1999). Both primer pairs were designed to anneal to evolutionarily conserved coding sequences corresponding to the evolutionary conserved TM3 (MAYDRY [ORFor1] and ICKPLHY [ORFor2]) and TM7 (PMLNPLIY [ORRev1] and NPFIYS(F/L) [ORRev2]) OR protein regions. Both primer pairs were predicted to generate products of ~ 0.5 kb representing approximately half the expected OR full-length coding sequence (vertebrate OR genes have an open reading frame of ~ 1kb which lacks introns (Mombaerts, 1999b).

Reverse transcriptase (RT)-PCR was carried out using a commercial kit (OneStep RT-PCR Kit, Qiagen) with the following thermocycling parameters: 94ºC/2 min; 94ºC/30 sec; 37ºC/30 sec, ramping from 37ºC to 72ºC at 0.2ºC/sec; 72ºC/60 sec; 5 cycles; 94ºC/30 sec; 45ºC /30 sec; 72ºC/60 sec; 30 cycles; 72ºC/7 min; 4ºC/hold.

As a control for the quality of the cDNA, chicken β-actin cDNA sequences were amplified using the primer pair: forward: 5'-GAGAAATTGTGCGTGACATCA-3' (corresponding to coordinates 3003 - 3023 of GenBank accession number **X00182**) and reverse: 5'-CCTGAACCTCTCATTGCCA-3' (corresponding to coordinates 3136- 3154 of GenBank accession no. **X00182**) with the following thermocycling parameters s: 94ºC/5min; 94ºC/30 sec, 50ºC/30 sec, 72ºC/1 min, 30 cycles; 72ºC/7 min; 4ºC/hold. The amplification product was predicted to be 152 bp.

Table 5.1
Percent sequence identities between the six chicken testis expressed OR sequences (GgTestesOR-A-F)

		% nucleotide identity					
	Sequence	GgTestesOR-A	GgTestesOR-B	GgTestesOR-C	GgTestesOR-D	GgTestesOR-E	GgTestesOR-F
	GgTestesOR-A		95	50	87	93	56
	GgTestesOR-B	92		50	91	97	58
% amino acid identity	GgTestesOR-C	36	38		50	50	51
	GgTestesOR-D	74	81	37		91	56
	GgTestesOR-E	88	95	38	81		58
	GgTestesOR-F	42	45	40	40	45	

Table 5.2
Results of BLAT searches of the red jungle fowl (*Gallus gallus*) genome sequence (Build 2.1) using the six OR partial coding sequences amplified from chicken testes cDNA. Chromosome numbers, location coordinates, corresponding RefSeq Genes, and the percent identity of the best alignments (highest score) are shown.

Sequence	Chromosome	location coordinates	RefSeq Gene	% identity of best alignment
GgTestesOR-A	5	1047016-1047492	DNAJC17 (Red jungle fowl)	95
GgTestesOR-B	5	1047016-1047492	DNAJC17 (Red jungle fowl)	100
GgTestesOR-C	10	109689-110165	Cor7a (Red jungle fowl)	99.8
GgTestesOR-D	5	1064636-1065112	Cor3b (Red jungle fowl)	94.8
GgTestesOR-E	5	1047016-1047492	DNAJC17 (Red jungle fowl)	97.3
GgTestesOR-F	5	1041625-1042101	OR7E5P (Human)	99.6

Chapter 6

Biogenic trace amine-associated receptors (TAARs) are encoded in avian genomes: evidence and possible implications

Jakob C. Mueller, Silke S. Steiger, Andrew E. Fidler and Bart Kempenaers

ABSTRACT

Recent studies of mammals and fish indicate that most trace amine-associated receptors (TAARs) may be involved in the detection of volatile biogenic compounds. It has therefore been suggested that this new class of 'olfactory' receptors could be highly relevant for social communication and individual recognition. To determine if TAAR orthologues are encoded in avian genomes, we initiated BLAST searches of the red jungle fowl (*Gallus gallus*) genome and public avian EST databases, and performed associated phylogenetic analyses of the TAAR homologues identified. Our results suggest that a minimum of three TAAR paralogues are encoded in the red jungle fowl genome, and that these are putative orthologues of the human/mouse genes TAAR1, TAAR2 and TAAR5. It is noteworthy that TAAR5 is activated by compounds that have been found in avian faeces. We tentatively suggest that avian TAARs may compensate for the lack of an avian equivalent of the mammalian vomeronasal system and therefore may be important mediators of socially important avian chemical cues.

INTRODUCTION

Recently, a new gene family of G-protein coupled receptors (GPCRs), denoted trace amine-associated receptors (TAARs), have been identified in both mammalian and fish genomes (Borowsky et al., 2001; Gloriam et al., 2005). Because trace amines are both structurally similar to conventional amine neurotransmitters and are present in neural tissue, albeit in small amounts, TAARs have been investigated for functions in neurotransmission and neuromodulation (Lindemann and Hoener, 2005; Lewin, 2006). Although some evidence for such functions has been reported, recent studies point to TAARs being mainly involved in olfaction (Liberles and Buck, 2006). Most TAAR genes are selectively expressed in small subsets of olfactory sensory neurons within the olfactory epithelium. In contrast to the large, and structurally diverse, odorant receptor family, which mediates general olfaction, TAARs appear to be specialised for detecting biogenic volatile amines. It has therefore been suggested that this new class of chemosensory TAAR receptors is associated with the detection of social cues and thus may be highly relevant for the study of behaviours associated with individual recognition and/or mate choice (Liberles and Buck, 2006). In light of growing evidence that birds do indeed have a well-developed sense of smell which may be of considerable behavioural significance, we sought to determine if avian genomes encode TAAR orthologues.

METHODS AND RESULTS

To determine if TAAR orthologues are encoded in avian genomes, we initiated a search for TAAR homologues in both the red jungle fowl (*Gallus gallus*) genome and additional public avian EST databases. The human TAAR1 protein sequence (RefSeq accession number NP_612200) was used as the query string for TBLASTN searches of the NCBI red jungle fowl reference genome database (Build 2.1, released in May 2006). To distinguish genuine TAAR orthologues from homologous non-TAAR GPCRs (e.g. serotonin, dopamine and adrenergic receptors), the cut-off E value for significant matches was set, on the basis of initial searches, at 10^{-50}. Using the NCBI default BLAST search settings TBLASTN searches returned four matches with E-values $<10^{-50}$ on a contig assigned to G. gallus chromosome 3 (accession number NW_001471670.1) with E-values of 10^{-132}, 10^{-87}, 10^{-77}, and 10^{-72}. However, the aligned sequence of the last match was substantially shorter (289 amino acids) than the first three matches (between 326 and 335 amino acids). There is no associated RefSeq gene annotation in the red jungle fowl genome, but sequence conservation with other vertebrate taxa appears to be high in the matched regions.

To obtain the full-length coding sequences of the putative red jungle fowl TAAR genes, the red jungle fowl genome assembly of May 2006 (UCSC genome browser) was searched for the longest open reading frames (ORF) corresponding to each of the putative genes, using the software package BioEdit (Hall, 1999). Three continuous ORFs with lengths of either 996 or 1026 bp were found. The fourth putative ORF was truncated by an assembly gap, however the 894 bp of sequence found was 100 % identical with one of the three other ORFs suggesting a duplication error during genome assembly. The lengths of the three unique putative TAAR ORFs (Figure 6.1) conformed to the expectation of intronless ORFs encoding 316–384 amino acid proteins (Gloriam et al., 2005). In addition, the predicted proteins share the TAAR-

specific peptide fingerprint motif (Figure 6.1) proposed to distinguish TAARs from other GPCRs (Lindemann and Hoener, 2005).
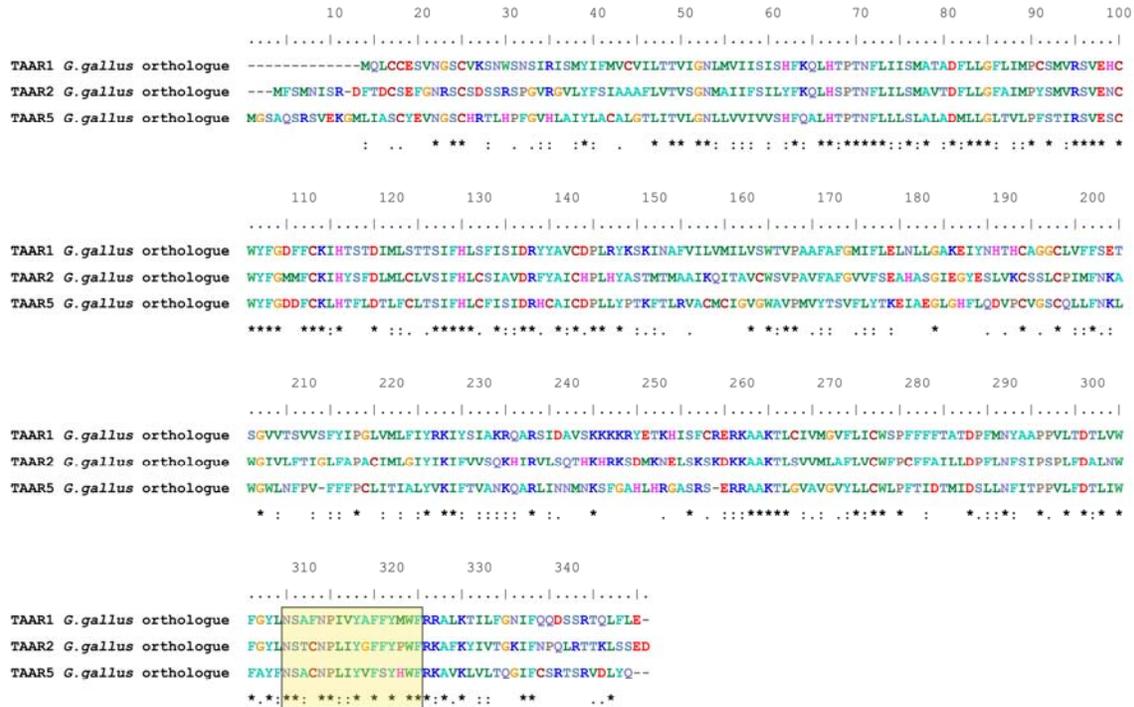
Figure 6.1
ClustalW alignment of three red jungle fowl (*Galllus gallus*) predicted TAAR protein sequences. A motif thought to define TAAR proteins is highlighted in yellow, while positions of identity in all sequences are indicated by asterisks.

The human, *Homo sapiens*, and mouse, *Mus musculus*, orthologues of the three putative red jungle fowl TAAR genes were determined by BLASTP searches of the NCBI RefSeq database of human and mouse protein sequences. The significantly best 'hits' for the putative red jungle fowl TAAR proteins were TAAR1 (human, NP_612200.1, E = 10^{-121}; mouse, NP_444435.1, E = 10^{-111}), TAAR2 (human, NP_001028252.1, E = 10^{-118}; mouse, NP_001007267.1, E = 10^{-124}) and TAAR5 (human, NP_003958.1, E = 10^{-119}; mouse, NP_001009574.1, E = 10^{-118}), with all three orthologues located on continuous regions of human chromosome 6 or mouse chromosome 10. These chromosomal assignments are consistent with previously reported synteny between large sections of red jungle fowl chromosome 3 with human chromosome 6 and mouse chromosome 10 (International chicken genome sequencing consortium 2004). In the red jungle fowl – human comparisons, the amino acid sequence identities of the putatively orthologous TAAR1, TAAR2 and TAAR5 proteins were 69%, 64% and 70%, respectively, and the equivalent chicken – mouse comparisons were 63%, 64% and 69%, respectively. These values of TAAR protein

identities lie between values reported for analogous fish – mammal comparisons (35 - 50%) and mammal - mammal comparisons (80-98%) (Gloriam et al., 2005). TBLASTN searches of the total avian EST database at NCBI did not return any alignments with E values $<10^{-37}$, indicating that no homologues of the G. gallus TAAR sequences were present in the avian EST databases.

A phylogenetic analysis using the coding sequences of all known mouse, human and red jungle fowl TAAR genes verified the homology relationships amongst the TAAR genes of these three vertebrates (Figure 6.2). The three red jungle fowl TAAR sequences are clearly located within the TAAR gene family and cluster within the TAAR1, TAAR2 and TAAR5 mammalian orthologues as expected.

DISCUSSION

Our results suggest that a minimum of three TAAR paralogues are encoded in the red jungle fowl genome, and that these are putative orthologues of the mammalian genes TAAR1, TAAR2 and TAAR5. Two of the three genes (TAAR1 and TAAR2) were detected, but are not explicitly described in any detail, in a report of the complete red jungle fowl GPCR family (Lagerstrom et al., 2006). Bearing in mind the current draft status of the red jungle fowl genome, our estimate of the number of red jungle fowl TAAR genes must be regarded as conservative and is at the lower end of the estimates described for other species: human, 6; mouse, 15; zebrafish, *Danio rerio*, 57; fugu, *Takifugu rubripes*, 8 (Gloriam et al., 2005; Liberles and Buck, 2006). Evidence from mouse and fish indicates that most TAAR genes, with the exception of TAAR1, are expressed in the olfactory epithelium (Liberles and Buck, 2006). The absence of TAAR homologous sequences from the avian EST databases is likely due to both the typically low levels of TAAR transcripts (Lindemann and Hoener, 2005; Lewin, 2006) and the paucity of cDNA libraries generated from avian olfactory epithelia. Assuming that red jungle fowl TAAR1, like its murine and piscine orthologues, is not expressed in the olfactory epithelium, we propose that the avian TAAR2 and TAAR5 orthologues are candidate avian receptors for the detection of volatile biogenic compounds. Clearly investigations of TAAR2/5 gene expression patterns in avian olfactory epithelium, and other tissues, as well as functional studies are necessary to confirm this hypothesis. It is noteworthy that TAAR5 is activated by compounds - specifically trimethylamine, N-methylpiperidine and isoamylamine - found in the urine of sexually mature male mice (Liberles and Buck, 2006). Interestingly, in the context of this report, volatile trimethylamine has been found in faeces of the black-bellied whistling duck (*Dendrocygna autumnalis*) (Robacker et al., 2000). We tentatively suggest that similar amino-acid derivatives, although perhaps less structurally diverse than in mammals, in interaction with their cognate TAAR receptors, may mediate avian social communication and/or individual recognition in some contexts. For example it has been shown that the Antarctic prion (*Pachiptila desolata*) is able to recognize its breeding partner by individual specific odours (Bonadonna and Nevitt, 2004). Furthermore, the significance of TAAR receptors in mediating social cues may be particularly pronounced in birds, as they seem to lack both the vomeronasal organ and vomeronasal receptors thought to mediate social chemo-communication in mammals (International Chicken Genome Sequencing Consortium, 2004).
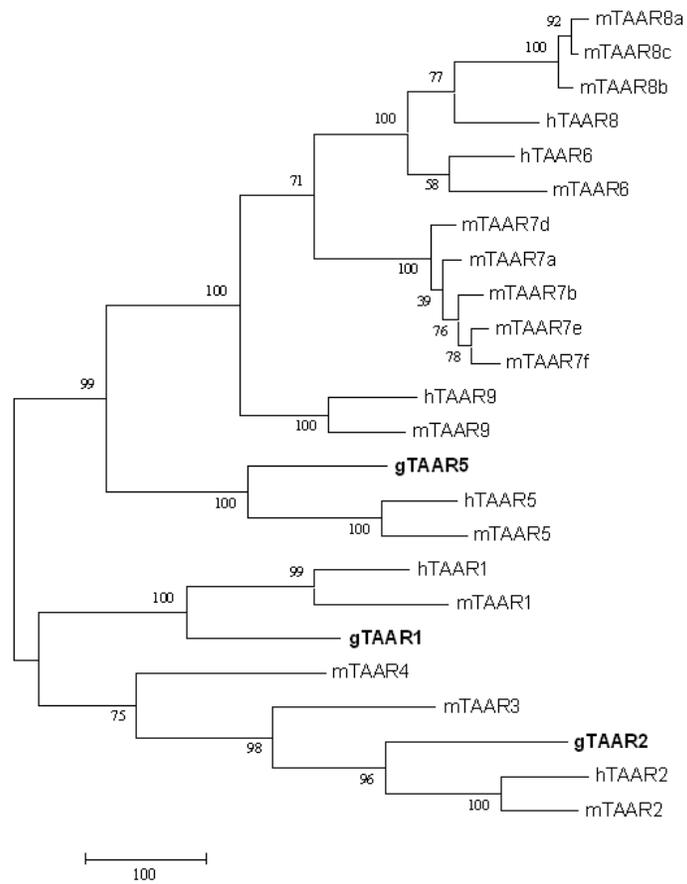
Figure 6.2
Maximum parsimony tree of all human, mouse and predicted red jungle fowl TAAR coding DNA sequences. The bootstrap percentages of 500 bootstrap replications are given at the nodes and the following prefixes were used: h = human, m = mouse, g = Gallus gallus in bold. The maximum parsimony analysis is based on the heuristic search algorithm close-neighbour-interchange as implemented by MEGA (Tamura et al., 2007).

Chapter 7

General discussion

The molecular basis of the sense of smell is mediated by olfactory receptors (ORs) expressed on sensory neurons in the olfactory epithelium (Buck and Axel, 1991). Avian OR genes have hardly been studied. Thus, the major objective of this thesis was to provide more insights into the genetic basis of the sense of smell in birds. I therefore investigated the OR gene repertoires (both the proportion of potentially functional and the total number of OR genes) in both distantly and closely related bird species (Chapter 2 and 3). In addition, I examined whether there is evidence for adaptive evolution in avian OR genes (Chapter 4). Further, I investigated whether OR gene transcripts can be detected in non-olfactory tissues (i.e. testes, Chapter 5). Finally, I examined whether avian trace amine-associated receptors (TAARs), a second class of chemosensory receptors in vertebrates, are encoded in avian genomes (Chapter 6). In this Chapter, I summarize and discuss the major findings of this thesis, and suggest some directions for future research.

A HIGH PROPORTION OF POTENTIALLY FUNCTIONAL OLFACTORY RECEPTOR GENES IS ENCODED IN AVIAN GENOMES

The proportion of potentially functional OR genes provides insights into the selective pressures that have acted on the OR genes (Rouquier et al., 2000; Niimura and Nei, 2006). If olfaction has become less important in the evolutionary past of the species, a relaxation of selection may have led to an increase in the number of pseudogenes (no strong selection against loss-of-function mutations). In contrast, if the sense of smell was relevant in the evolutionary past of the species, selection may have maintained the functionality of OR genes.

Using polymerase chain reaction (PCR) with degenerate primers specific for the OR gene family that anneal to evolutionary conserved protein regions, I identified partial coding OR sequences 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 petrel, *Pagodroma nivea*) (Chapter 2). Primers were designed to anneal to transmembrane domain (TM) 3 and TM7 regions of the OR proteins and were predicted to generate products of approximately half of the expected full OR coding sequence. Interestingly, the large majority of the partial OR coding sequences was potentially functional (i.e. no stop codon) in all species investigated (on average 83.7% ± 2.3%). This result was supported when I estimated the proportion of potentially functional OR genes in two sets of closely related bird species (Paleognath genomes: brown kiwi, *Apteryx australis*; emu, *Dromaius novaehollandiae*; rhea, *Rhea americana*; ostrich, *Struthio camelus*; Psittaciform genomes: kakapo, *Strigops habroptilus*; kaka, *Nestor meridionalis*; kea, *Nestor notabilis*) (Chapter 3). Thus, the results suggest that there is selection to maintain the functionality of avian OR genes.

The results presented in Chapter 2 and 3 were surprising, because a previous analysis of the first draft of the red jungle fowl genome sequence (Build 1.1, released in February 2004) led to the conclusion that the chicken OR repertoire consisted largely of pseudogenes (Niimura and Nei, 2005) (Table 7.1). However, we now have evidence that Niimura and Nei's (2005) estimate of the proportion of functional OR genes based on the first draft of the red jungle fowl genome was an underestimate (Kuryshev, Steiger, Stensmyr, Kempenaers and Müller, in preparation).

106

We conducted a database search for OR genes in the second, higher quality draft of the red jungle fowl genome (Build 2.1, released in May 2006) and found that a minimum of 53% are potentially functional (Table 7.1). Moreover, it is likely that many of the OR genes that are now categorized as partial genes will turn out to be functional when an improved version of the red jungle fowl genome will become available. This is because many of the partial genes still contain gaps within the sequence (130 out of 192; Table 7.1). Under the assumption that all partial OR genes will turn out to be functional, we estimate that a maximum of 85% of the red jungle fowl OR genes is functional. Note that this estimate is very similar to the estimate we obtained with the degenerate PCR approach (Table 7.1). In conclusion, the PCR based approach seems to be useful and legitimate to investigate OR gene repertoires in species for which full genomic sequences are not yet available.

Table 7.1
The number and proportion of functional olfactory receptor (OR) genes identified in the red jungle fowl (*Gallus gallus*) genome. Abbreviation: na, not applicable.

Method	Red jungle fowl genome draft	Total number of OR genes (functional/pseudogenes/partial)	Percentage of functional OR genes	Reference
Database Mining	Build 1.1 [1]	283 (na/na/na)	na	International Chicken Genome Sequencing Consortium (2004)
Database Mining	Build 1.1 [1]	558 (82/476[3])	15	Niimura and Nei (2005)
Database Mining	Build 1.1 [1]	na (229/na/na)	na	Lagerstrom et al. (2006)
Database Mining	Build 2.1 [2]	406 (214/62/130)	53	Kuryshev et al., in preparation
PCR using degenerate primers		638	96	Steiger et al., submitted; Chapter 2

[1] released in February 2004
[2] released in May 2006
[3] both pseudogenes and partial OR genes

THE ESTIMATED NUMBERS OF OLFACTORY RECEPTOR GENES ENCODED IN AVIAN GENOMES VARY WIDELY AMONG SPECIES

Previous studied have identified entire OR gene repertoires in species for which full genomic sequences are available (Glusman et al., 2001; Young et al., 2002; Gimelbrant et al., 2004; International Chicken Genome Sequencing Consortium, 2004; Alioto and Ngai, 2005; Niimura and Nei, 2005; Quignon et al., 2005; Warren et al., 2008). So far, the red jungle fowl genome is the only bird genome available and – with the exception of the zebra finch, *Taeniopygia guttata*, genome – it is highly unlikely that many bird genomes

will be sequenced in the near future. Thus, a nonparametric estimation technique applying the concept of 'sample coverage' (Chao and Lee, 1992) was used to estimate the total number of OR genes in a wide range of avian genomes. This method has been used in other contexts (Huang and Weir, 2001), e.g. to estimate the total number of alleles in a population, and has been shown to perform better than other methods (Bunge and Fitzpatrick, 1993). It is important to note that the nonparametric estimation technique does not assume an equal probability for each gene to be cloned and thus accounts for primer bias. Our estimate for the red jungle fowl, 638 OR genes, is close to the previously reported estimate of 558 derived by Niimura and Nei (2005) (Table 7.1), suggesting that our methodology provides a reasonably reliable estimate of OR gene repertoire size. However, also note that there is a substantial disagreement in the estimated number of OR genes, probably because different bioinformatics search strategies were used (Table 7.1).

To my knowledge, only one other study exists that used a statistical method to estimate OR gene repertoire sizes. In this study (Glusman et al., 2000), the distribution of ORs that appeared n times as obtained experimentally was fitted to a set of binomial distributions. Primer bias was not taken into account, because an equal probability for each gene to be cloned was assumed. The OR gene repertoire from a marsupial (koala, *Phascolarctos cinereus*) and a monotreme (platypus, *Ornithorhynchus anatinus*) was estimated to be extremely small (~50 OR genes) (Glusman et al., 2000). Yet, large numbers of ORs (~ 700) have been recently detected in the platypus genome assembly (Warren et al., 2008). Therefore, Glusman et al.'s estimation approach did not yield a reliable estimate of the platypus OR gene repertoire size.

The estimated numbers of OR genes encoded in avian genomes varied up to 6-fold between species (range: 106 – 667 OR genes in the galah and the kakapo, respectively) (Chapter 2), indicating that different ecological niches may have shaped the OR gene repertoires in birds, as has been suggested for mammals (Niimura and Nei, 2007). The observed differences in OR gene repertoire sizes are striking, but perhaps not too surprising as birds also show wide (up to 12-fold) interspecific variation in the relative olfactory bulb size, as quantified by the olfactory bulb ratio (OBR). Hence, similar interspecific variation in OR gene repertoire size could be expected. As the total number of OR genes in a genome may reflect how many different scents can be detected and distinguished (Niimura and Nei, 2006), these results suggest an excellent developed sense of smell in bird species with a large OR gene repertoire.

Interestingly, nocturnal species that rely on their sense of smell but that lack a well-developed sense of vision (the brown kiwi and the kakapo (Wenzel, 1968; Hagelin, 2004; Martin et al., 2007) encoded a larger number of OR genes than their diurnal counterparts, suggesting that ecological factors such as activity patterns contributed to the evolution of avian OR gene repertoires (Chapter 3). The large number of estimated OR genes in the two nocturnal species may contribute to the ability of these species to locate food at night via olfactory cues (Wenzel, 1968; Hagelin, 2004). We estimated that the kakapo and the kiwi have ~ 450-550 functional OR genes. Thus, their sense of smell may be at least as good as the sense of smell in humans (the human OR gene repertoire is estimated to encode 339 functional OR genes; Malnic et al., 2004).

THE ESTIMATED NUMBERS OF OLFACTORY RECEPTOR GENES ARE POSITIVELY CORRELATED WITH THE RELATIVE SIZE OF THE OLFACTORY BULB

The olfactory bulb ratio (OBR), an anatomical correlate of olfactory capability (Edinger, 1908), is the ratio of the greatest diameter of the olfactory bulb to the greatest diameter of the cerebral hemisphere in per cent (Bang and Cobb, 1968) In Chapter 2, I showed 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 the OBR. It should be noted that the correlation was only weakly significant and not significant anymore when a more recent tree topology (Cracraft et al., 2004) was used to control for phylogeny. Yet, the positive correlation was strongly supported in Chapter 3, in which the OR gene repertoire sizes of two sets of closely related species were compared (Paleognath genomes: brown kiwi, *Apteryx australis*; emu, *Dromaius novaehollandiae*; rhea, *Rhea americana*; ostrich, *Struthio camelus*; Psittaciform genomes: kakapo, *Strigops habroptilus*; kaka, *Nestor meridionalis*; kea, *Nestor notabilis*). Here, the correlation between the estimated number of OR genes and the OBR was significant despite the use of a more recent tree topology.

ALL BIRD GENOMES EXAMINED ENCODE OLFACTORY RECEPTOR GENES BELONGING TO A LARGE GENE CLADE, TERMED γ-C

Phylogenetic Neighbour-Joining (NJ) trees derived from predicted OR protein sequences revealed that all examined bird genomes encoded OR genes belonging to a large gene clade, termed group-γ-c, whose expansion appears to be a shared characteristic of class Aves (Chapter 2). Interestingly, within the γ-c clade, there is a strong tendency for sequences from the same species, or species from the same order, to cluster together, suggesting 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. In contrast, amongst 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. As the group-γ-c genes cluster species-specifically, it is likely that these ORs may detect chemical stimuli uniquely encountered by each species. Two lines of evidence suggest that the γ-c clade expansion did not occur earlier in evolution. Firstly, we used degenerate primer pairs to amplify OR genes from the Nile crocodile, *Crocodylus niloticus*. No group-γ-c OR genes could be amplified with these primer pairs. Secondly, within the scope of a comparative study (comparison of anolis lizard, *Anolis carolinensis*, and red jungle fowl OR genes), we did not detect any group-γ-c OR genes in the anolis lizard genome (Kuryshev et al., in preparation; also see Chapter 2 and Figure 7.1 – 7.2). It would be worthwhile to investigate the functional significance and the ligands of the apparently bird lineage specific expanded γ-c OR clade.

THERE IS EVIDENCE FOR POSITIVE SELECTION ON AVIAN OLFACTORY RECEPTOR GENES

Positive selection, i.e. selection favouring changes in the protein sequence, plays a role in the evolution of OR gene repertoires in fish (Ngai et al., 1993; Kondo et al., 2002; Alioto and Ngai, 2005) and mammals (Hughes and Hughes, 1993; Singer et al., 1996; Gilad et al., 2000). I used three different maximum likelihood methods to identify positively

selected codons in partial, group-γ-c OR coding sequences in the same nine bird species mentioned above (Chapter 4). In addition, I tested for alignment-wide inference of selection.

I showed that there is evidence for positive selection at particular codons in every bird species studied. Interestingly, positively selected codon sites were predominantly located in transmembrane domain (TM) regions (i.e. TM4, TM5 and TM6) that are most likely involved in odorant binding in other vertebrate species (Ngai et al., 1993; Hall et al., 2004; Alioto and Ngai, 2005; Katada et al., 2005). It is reasonable to assume that the positively selected codons are functionally relevant, because even a single point mutation in the binding site of an OR can alter the ligand specificity thereby allowing the OR to recognize certain odorant molecules with higher affinities and others with lower affinities (Katada et al., 2005).

Strikingly, both alignment-wide evidence of positive selection and a large number of positively selected codon sites were detected in the brown kiwi, a bird known to have an excellent sense of smell (Wenzel, 1968; Martin et al., 2007). The number of positively selected sites correlated positively with both the OBR and the estimated number of OR genes encoded in the bird genomes, suggesting that the extent of adaptive evolution is higher in bird species that have good olfactory abilities. Further work could compare evidence for positive selection in different OR gene clades (γ-c, non-γ-c and group-α).

Figure 7.1 (opposite page)
(a) Unrooted Neighbour-Joining (NJ) tree generated from an alignment of red jungle fowl (*Gallus gallus*, pink lines) and anolis lizard (*Anolis carolinensis*, black lines) predicted full-length OR protein sequences. The red jungle fowl and anolis lizard predicted OR protein sequences were obtained from Kuryshev et al., unpublished data. The group-γ-c OR genes are indicated with an arrow. The scale-bar indicates the number of amino acid substitutions per site.   (b) Unrooted Neighbour-Joining (NJ) tree generated from an alignment of western-clawed frog (*Xenopus tropicalis*, brown lines), zebrafish (*Danio rerio,* light green lines), pufferfish (*Fugu rubripes*, dark green lines), red jungle fowl (*Gallus gallus*, blue lines), anolis lizard (*Anolis carolinensis*, red lines) and human (*Homo sapiens*, black lines) predicted full-length OR protein sequences. The red jungle fowl and anolis predicted OR protein sequences were obtained from Kuryshev et al. (in preparation), whereas the other OR protein sequences were obtained from Niimura and Nei (2005). The group-γ-c OR genes are indicated with an arrow. The scale-bar indicates the number of amino acid substitutions per site.
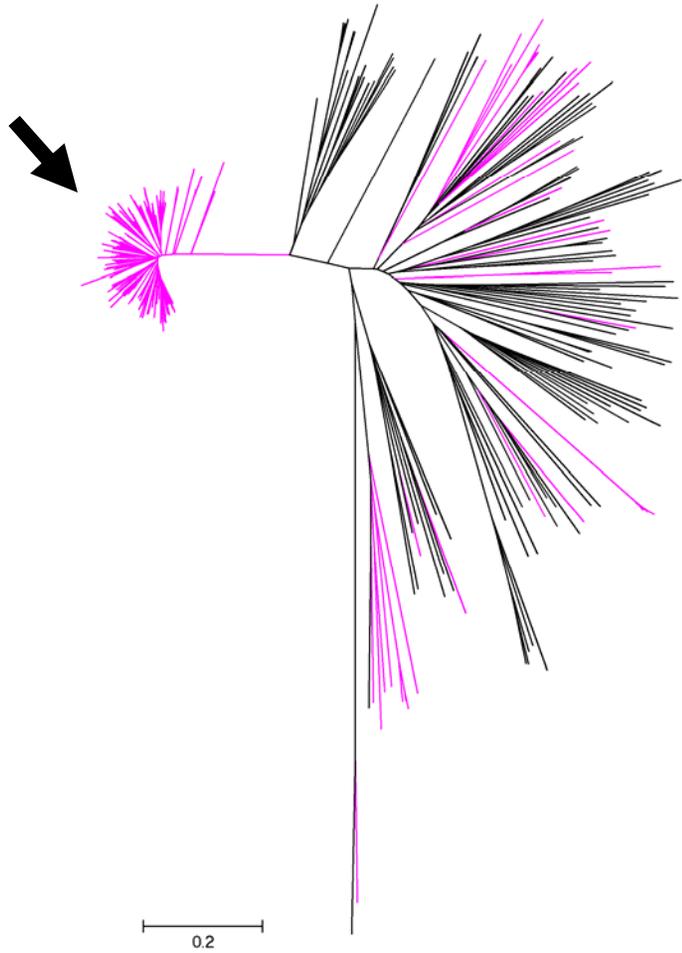
(a)
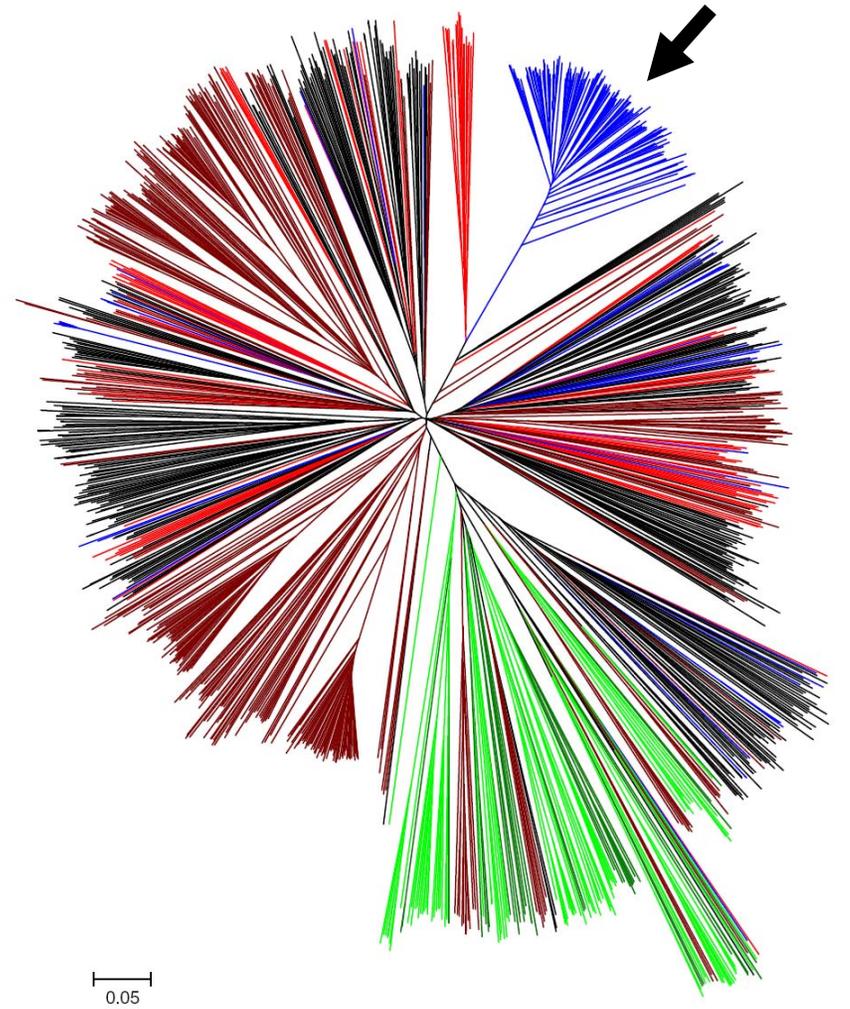
(b)

OLFACTORY RECEPTOR TRANSCRIPTS CAN BE DETECTED IN BIRD TESTES

OR gene expression is not restricted to the olfactory tissue. Interestingly, transcripts of a subset of OR genes have been detected in testis and sperm (e.g. Parmentier et al. 1992). More specifically, mammalian testicular/sperm OR proteins have been localized in the sperm flagellum midpiece where they may mediate flagellar motility (Spehr et al., 2003). Hence, sperm expressed OR genes may play an important role in sperm-egg chemotaxis/communication (Spehr et al., 2004).

Using reverse transcription (RT)-PCR with degenerate primers specific for OR genes, and subsequent cloning, I showed that a minimum of six OR gene transcripts are present in chicken (*Gallus gallus domesticus*) testes (Chapter 5). The OR transcripts amplified can be placed in the group-γ, but not in the group-γ-c subclade, whose expansion is apparently specific to bird genomes. I detected no premature stop codons within the predicted open reading frames in any of the OR partial coding sequences, suggesting that all sequences were derived from transcripts encoding functional OR proteins. It remains to be tested whether avian testicular OR proteins can been localized to the sperm flagellum midpiece and will mediate flagellar motility (Spehr et al., 2003).

TAAR GENES ARE ENCODED IN AVIAN GENOMES

Trace amine-associated receptors (TAARs) are a second class of G-protein coupled receptors (GPCRs) expressed in olfactory neurons in the mouse (Gloriam et al., 2005; Liberles and Buck, 2006). This suggests that they play a role as chemosensory receptors in vertebrates. BLAST searches of the red jungle fowl genome and public avian EST databases were used to show that a minimum of three TAAR genes are encoded in the red jungle fowl genome (Chapter 6). A phylogenetic analysis of the identified TAAR homologues revealed that they are putative orthologues of the human/mouse genes TAAR1, TAAR2 and TAAR5.

The TAAR gene family repertoire varies substantially among vertebrates. It has recently been shown that fish have many functional TAAR genes (range: 13-109), whereas tetrapods encode fewer TAAR genes (range: 3-22) (Hashiguchi and Nishida, 2007) (Table 7.2). The red jungle fowl has the lowest estimated number of TAAR genes (Table 7.2) in vertebrates, suggesting that these chemosensory receptors may be of minor importance in birds. On the other hand, avian TAARs may compensate for the lack of an avian equivalent of the mammalian vomeronasal system and therefore may be important mediators of socially important avian chemical cues. Testing whether TAAR genes are expressed in the olfactory epithelium would help to understand whether TAAR genes are functionally relevant in birds.

Preliminary evidence indicates that TAAR genes are also encoded in at least two other bird genomes: Southern Blot analysis using TAAR5 as probe revealed one band in genomic digested DNA in the kakapo and the kiwi (Figure 7.3). Note that only one band is observed in the digested DNA of both species, indicating that homologues of the TAAR5 gene most likely occur in relatively small numbers. Whether TAAR1 and

TAAR2 homologues are encoded in genomes of bird species other than the red jungle fowl awaits further study.

Table 7.2
Trace amine-associated receptor (TAAR) genes in vertebrates (adapted from Hashiguchi and Nishida (2007) and Müller et al. (2008)).

Common name	Latin name	Number of TAAR genes (functional/pseudogenes)
Jawless fish:		
Sea lamprey	*Petromyzon marinus*	38 (21/17)
Teleost fish:		
Zebrafish	*Danio rerio*	119 (109/10)
Medaka	*Oryzias latipes*	32 (25/7)
Stickleback	*Gasterosteus aculeatus*	64 (49/15)
Fugu	*Takifugu rubripes*	19 (13/6)
Amphibians:		
Western clawed Frog	*Xenopus tropicalis*	7 (6/1)
Birds:		
Red jungle fowl	*Gallus gallus*	5 (3/2)
Mammals:		
Opossum	*Monodelphis domestica*	35 (22/13)
Mouse	*Mus musculus*	16 (15/1)
Human	*Homo sapiens*	8 (5/3)

FUTURE DIRECTIONS

The draft genome sequence of a passerine bird, the zebra finch, will be released in the near future. An extended comparative study that encompasses complete OR gene repertoires of both the red jungle fowl and the zebra finch will shed further light on the evolutionary dynamics of avian OR genes. For example, selective pressures that have been shaping the OR gene repertoire since the divergence of zebra finch and chicken could be studied. In addition, phylogenetic analysis could reveal the OR gene repertoire size of the ancestral species and allow to quantify gain and loss of genes that occurred in each evolutionary lineage (Nozawa and Nei, 2007).

So far, only few red jungle fowl OR genes have been localized, i.e. mapped onto chromosomes. A database search shows that the majority of red jungle fowl OR genes, in particular the group-γ-c OR genes, are localized on 'unknown chromosome', suggesting that OR loci may be rich in repetitive DNA sequences and may thus be difficult to map (Zhang et al., 2004b). As knowledge of the chromosome localization pattern is important for our understanding of how the sequence diversity of OR genes was generated, future studies could examine whether avian OR genes occur in clusters on chromosomes, as has been shown for other vertebrate species (Sharon et al., 1999; Glusman et al., 2001; Niimura and Nei, 2003).
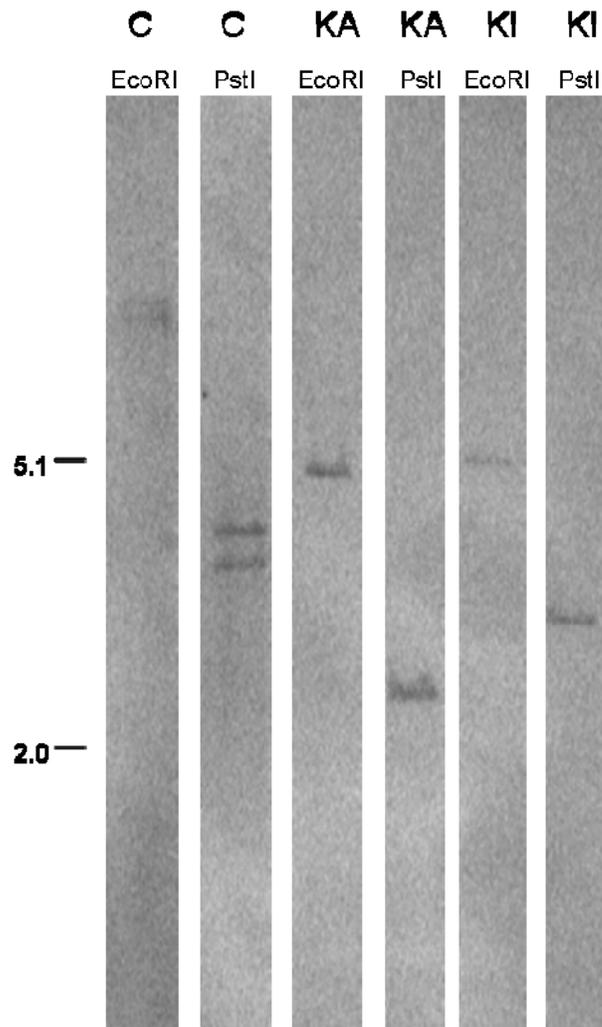
Figure 7.3
Southern hybridization of restriction enzyme digested chicken (*Gallus gallus domesticus*), kakapo (*Strigops habroptilus*) and brown kiwi (*Apteryx australis*) genomic DNA. Genomic DNA was digested with two different restriction enzymes (EcoRI and PstI) and used for Southern hybridization with a digoxigenin (DIG)-labeled probe generated from the chicken TAAR5 gene. High stringency hybridization conditions were used (42°C hybridization temperature). Abbreviations: C, Chicken; KA, Kakapo; KI, Kiwi. Approximate positions of the size standards (in kb) are indicated.

So far, expression studies of OR genes have hardly been conducted - even in mammals (Zhang et al., 2004a; Feldmesser et al., 2006; Zhang et al., 2007). Characterizing expression levels of avian OR genes would allow to investigate the degree to which OR genes are expressed in olfactory and non-olfactory tissues. For example, custom-made microarrays including all known mouse olfactory receptor (OR) genes were used to show specific expression in the olfactory sensory epithelium of more than 800 OR genes (Zhang et al., 2004a). Previously, these genes were designated as ORs, based entirely on genomic sequences. A similar approach with the red jungle fowl OR genes could be used to tackle the question how many OR genes are expressed in the bird olfactory epithelium. In addition, this method could be used to determine whether the same OR genes expressed in the olfactory epithelium are also expressed in non-olfactory tissues (Zhang et al., 2004a).

CONCLUSION

In summary, this thesis contributes to our understanding of the patterns and mechanisms of the evolution of avian OR genes. The molecular data presented in this thesis, in particular the estimated OR gene repertoire sizes and the proportion of OR genes that is potentially functional, suggest that olfaction may be more important for birds than is often acknowledged. Further studies of avian OR genes are needed to gain more detailed insights into the mechanisms and evolution of avian chemoreception.

REFERENCES

A

Ache, B. W. and Young, J. M. (2005). Olfaction: Diverse species, conserved principles. *Neuron* **48**, 417-430.

Ainley, D. G., O'Connor, E. F. and Boekelheide, R. J. (1984). The marine ecology of birds in the Ross Sea, Antarctica. *Ornithological Monographs* **32**, 1-97.

Alioto, T. S. and Ngai, J. (2005). The odorant receptor repertoire of teleost fish. *BMC Genomics* **6**, 173.

Amo, L., Galvan, I., Tomas, G. and Sanz, J. J. (2008). Predator odour recognition and avoidance in a songbird. *Functional Ecology* **22**, 289-293.

Anisimova, M. and Liberles, D. A. (2007). The quest for natural selection in the age of comparative genomics. *Heredity* **99**, 567-579.

Anisimova, M., Nielsen, R. and Yang, Z. H. (2003). Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229-1236.

B

Ball, G. F. and Hulse, S. H. (1998). Birdsong. *American Psychologist* **53**, 37-58.

Bang, B. G. (1971). Functional anatomy of the olfactory system in 23 orders of birds. *Acta Anatomica (Supplement)* **79**, 1-76.

Bang, B. G. and Cobb, S. (1968). Size of the olfactory bulb in 108 species of birds. *Auk* **85**, 55-61.

Bang, B. G. and Wenzel, B. M. (1985). Nasal cavity and olfactory system. In *Form and Function in Birds*, vol. 3 (ed. S. King and J.McClelland), pp. 195-225. New York Academic Press.

Birkhead, T. R., Martinez, J. G., Burke, T. and Froman, D. P. (1999). Sperm mobility determines the outcome of sperm competition in the domestic fowl. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 1759-1764.

Bonadonna, F. and Bretagnolle, V. (2002). Smelling home: a good solution for burrow-finding in nocturnal petrels? *Journal of Experimental Biology* **205**, 2519-2523.

Bonadonna, F. and Nevitt, G. A. (2004). Partner-specific odor recognition in an antarctic seabird. *Science* **306**, 835.

Borowsky, B., Adham, N., Jones, K. A., Raddatz, R., Artymyshyn, R., Ogozalek, K. L., Durkin, M. M., Lakhlani, P. P., Bonini, J. A., Pathirana, S. et al. (2001). Trace amines: Identification of a family of mammalian G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8966-8971.

Branscomb, A., Seger, J. and White, R. L. (2000). Evolution of odorant receptors expressed in mammalian testes. *Genetics* **156**, 785-797.

Buck, L. and Axel, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65**, 175-187.

Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species - a review. *Journal of the American Statistical Association* **88**, 364-373.

C

Chao, A. and Lee, S. M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* **87**, 210-217.

Chen, J.-M., Cooper, D. N., Chuzhanova, N., Ferec, C. and Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* **8**, 762-775.

Christidis, L. and Boles, W. E. (2008). Systematics and taxonomy of australian birds. Collingwood, Australia: Csiro.

Clark, L. and Mason, J. R. (1987). Olfactory Discrimination of Plant Volatiles by the European Starling. *Animal Behaviour* **35**, 227-235.

Clark, L. and Smeraski, C. A. (1990). Seasonal shifts in odor acuity by starlings. *Journal of Experimental Zoology* **255**, 22-29.

Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J. and Ward, R. (2001). Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.

Cracraft, J., Barker, F. K., Braun, M., Harshman, J., Dyke, G. J., Feinstein, J., Stanley, S., Cibois, A., Schikler, P., Beresford, P. et al. (2004). Phylogenetic relationships among modern birds (Neornithes). Toward an avian tree of life. In *Assembling the tree of life*, (ed. J. D. Cracraft and M.J. Donoghue), pp. 468-489. Oxford: Oxford University Press.

Cramp, S. and Perrins, C. M. (1994). Handbook of the birds of Europe, the Middle East and North Africa: The birds of the Western Palearctic: Crows to finches. Oxford: Oxford University Press.

Cuthill, I. C., Partridge, J. C., Bennett, A. T. D., Church, S. C., Hart, N. S. and Hunt, S. (2000). Ultraviolet vision in birds. *Advances in the Study of Behavior* **29**, 159-214.

D

Dahanukar, A., Hallem, E. A. and Carlson, J. R. (2005). Insect chemoreception. *Current Opinion in Neurobiology* **15**, 423-430.

de Kloet, R. S. and de Kloet, S. R. (2005). The evolution of the spindlin gene in birds: Sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes. *Molecular Phylogenetics and Evolution* **36**, 706-721.

del Hoyo, J., Elliot, A. and Christie, D. A. (2007). Handbook of the birds of the world: Picathartes to tits and chickadees. Barcelona, Spain: Lynx Edicions.

del Hoyo, J., Elliott, A. and Sargatal, J. (1992). Handbook of the birds of the world: Ostrich to ducks. Barcelona, Spain: Lynx Edicions.

E

Edinger, L. (1908). The relations of comparative anatomy to comparative psychology. *Journal of Comparative Neurology and Psychology* **18**, 437-457.

Elliot, A., Sargatal, J. and del Hoyo, J. (1994). Handbook of the birds of the world: New world vultures to guineafowl. Barcelona, Spain: Lynx Edicions.

Eriksson J., Larson G., Gunnarsson U., Bed'hom B., Tixier-Boichard M., Stromstedt L., Wright D., Jungerius A., Vereijken A., Randi E. et al. (2008).

Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *Plos Genetics* **4**, e1000010.

F

Feldmesser, E., Olender, T., Khen, M., Yanai, I., Ophir, R. and Lancet, D. (2006). Widespread ectopic expression of olfactory receptor genes. *BMC Genomics* **7**, 121.

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist* **125**, 1-15.

Fillon, V., Zoorob, R., Yerle, M., Auffray, C. and Vignal, A. (1996). Mapping of the genetically independent chicken major histocompatibility complexes B-@ and RFP-Y-@ to the same microchromosome by two-color fluorescent in situ hybridization. *Cytogenetics and Cell Genetics* **75**, 7-9.

Firestein, S. (2001). How the olfactory system makes sense of scents. *Nature* **413**, 211-218.

Floriano, W. B., Vaidehi, N. and Goddard, W. A. (2004). Making sense of olfaction through predictions of the 3-D structure and function of olfactory receptors. *Chemical Senses* **29**, 269-290.

Freitag, J., Beck, A., Ludwig, G., von Buchholtz, L. and Breer, H. (1999). On the origin of the olfactory receptor family: receptor genes of the jawless fish (Lampetra fluviatilis). *Gene* **226**, 165-174.

Freitag, J., Krieger, J., Strotmann, J. and Breer, H. (1995). Two classes of olfactory receptors in Xenopus laevis. *Neuron* **15**, 1383-1392.

Freitag, J., Ludwig, G., Andreini, I., Rossler, P. and Breer, H. (1998). Olfactory receptors in aquatic and terrestrial vertebrates. *Journal of Comparative Physiology A - Neuroethology Sensory Neural and Behavioral Physiology* **183**, 635-650.

Fuchs, T., Glusman, G., Horn-Saban, S., Lancet, D. and Pilpel, Y. (2001). The human olfactory subgenome: from sequence to structure and evolution. *Human Genetics* **108**, 1-13.

G

Gaillard, I., Rouquier, S. and Giorgi, D. (2004). Olfactory receptors. *Cellular and Molecular Life Sciences* **61**, 456-469.

Gilad, Y., Bustamante, C. D., Lancet, D. and Paabo, S. (2003). Natural selection on the olfactory receptor gene family in humans and chimpanzees. *American Journal of Human Genetics* **73**, 489-501.

Gilad, Y., Segre, D., Skorecki, K., Nachman, M. W., Lancet, D. and Sharon, D. (2000). Dichotomy of single-nucleotide polymorphism haplotypes in olfactory receptor genes and pseudogenes. *Nature Genetics* **26**, 221-224.

Gilad, Y., Wiebel, V., Przeworski, M., Lancet, D. and Paabo, S. (2004). Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *Plos Biology* **2**, 120-125.

Gimelbrant, A. A., Skaletsky, H. and Chess, A. (2004). Selective pressures on the olfactory receptor repertoire since the human-chimpanzee divergence. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9019-9022.

Gloriam, D. E. I., Bjarnadottir, T. K., Yan, Y.-L., Postlethwait, J. H., Schioth, H. B. and Fredriksson, R. (2005). The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish. *Molecular Phylogenetics and Evolution* **35**, 470-482.

Glusman, G., Bahar, A., Sharon, D., Pilpel, Y., White, J. and Lancet, D. (2000). The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mammalian Genome* **11**, 1016-1023.

Glusman, G., Yanai, I., Rubin, I. and Lancet, D. (2001). The complete human olfactory subgenome. *Genome Research* **11**, 685-702.

Godfrey, P. A., Malnic, B. and Buck, L. B. (2004). The mouse olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2156-2161.

Gwinner, H. and Berger, S. (2008). Starling males select green nest material by olfaction using experience-independent and experience-dependent cues. *Animal Behaviour* **75**, 971-976.

H

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.

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, N.H.: Science Publishers.

Hagelin, J. C. and Jones, I. L. (2007). Bird odors and other chemical substances: a defense mechanism or overlooked mode of intraspecific communication? *The Auk* **124**, 1–21.

Hall, S. E., Floriano, W. B., Vaidehi, N. and Goddard, W. A. (2004). Predicted 3-D structures for mouse 17 and rat 17 olfactory receptors and comparison of predicted odor recognition profiles with experiment. *Chemical Senses* **29**, 595-616.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Hashiguchi, Y. and Nishida, M. (2007). Evolution of trace amine-associated receptor (TAAR) gene family in vertebrates: Lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium. *Molecular Biology and Evolution* **24**, 2099-2107.

Healy, S. and Guilford, T. (1990). Olfactory-bulb size and nocturnality in birds. *Evolution* **44**, 339-346.

Huang, S. P. and Weir, B. S. (2001). Estimating the total number of alleles using a sample coverage method. *Genetics* **159**, 1365-1373.

Hughes, A. L. and Hughes, M. K. (1993). Adaptive evolution in the rat olfactory receptor gene family. *Journal of Molecular Evolution* **36**, 249-254.

I

        International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

        Irie-Kushiyama, S., Asano-Miyoshi, M., Suda, T., Abe, K. and Emori, Y. (2004). Identification of 24 genes and two pseudogenes coding for olfactory receptors in Japanese loach, classified into four subfamilies: a putative evolutionary process for fish olfactory receptor genes by comprehensive phylogenetic analysis. *Gene* **325**, 123-135.

        Issel-Tarver, L. and Rine, J. (1996). Organization and expression of canine olfactory receptor genes. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 10897-10902.

        Issel-Tarver, L. and Rine, J. (1997). The evolution of mammalian olfactory receptor genes. *Genetics* **145**, 185-195.

K

        Katada, S., Hirokawa, T., Oka, Y., Suwa, M. and Touhara, K. (2005). Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *Journal of Neuroscience* **25**, 1806-1815.

        Kishida, T., Kubota, S., Shirayama, Y. and Fukami, H. (2007). The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biology Letters* **3**, 428-430.

        Koenig, C. and Becking, J. H. (1999). Owls: A guide to the owls of the world. New Haven, CT Yale University Press.

        Kondo, R., Kaneko, S., Sun, H., Sakaizumi, M. and Chigusa, S. I. (2002). Diversification of olfactory receptor genes in the japanese medaka fish, Oryzias latipes. *Gene* **282**, 113-120.

        Konishi, M. and Knudsen, E. I. (1979). Oilbird - hearing and echolocation. *Science* **204**, 425-427.

        Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. and Frost, S. D. W. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096-3098.

        Kubo, Y., Akiyama, M., Fukada, Y. and Okano, T. (2006). Molecular cloning, mRNA expression, and immunocytochemical localization of a putative blue-light photoreceptor CRY4 in the chicken pineal gland. *Journal of Neurochemistry* **97**, 1155-1165.

        Kumar, S., Tamura, K. and Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.

L

        Lagerstrom, M. C., Hellstrom, A. R., Gloriam, D. E., Larsson, T. P., Schioth, H. B. and Fredriksson, R. (2006). The G protein - coupled receptor subset of the chicken genome. *PloS Computational Biology* **2**, 493-507.

        Lancet, D. and Ben-Arie, N. (1993). Olfactory receptors. *Current Biology* **3**, 668-674.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Laska, M., Genzel, D. and Wieser, A. (2005). The number of functional olfactory receptor genes and the relative size of olfactory brain structures are poor predictors of olfactory discrimination performance with enantiomers. *Chemical Senses* **30**, 171-175.

Leibovici, M., Lapointe, F., Aletta, P. and Ayer-Le Lievre, C. (1996). Avian olfactory receptors: Differentiation of olfactory neurons under normal and experimental conditions. *Developmental Biology* **175**, 118-131.

Lewin, A. H. (2006). Receptors of mammalian trace amines. *Aaps Journal* **8**, E138-E145.

Liberles, S. D. and Buck, L. B. (2006). A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645-650.

Lindemann, L. and Hoener, M. C. (2005). A renaissance in trace amines inspired by a novel GPCR family. *Trends in Pharmacological Sciences* **26**, 274-281.

Liu, A. H., Zhang, X. M., Stolovitzky, G. A., Califano, A. and Firestein, S. J. (2003). Motif-based construction of a functional map for mammalian olfactory receptors. *Genomics* **81**, 443-456.

M

Maddison, W. P. and Maddison, D. R. (2006). Mesquite: A modular system for evolutionary analysis. Version 1.12. See http://mesquiteproject.org.

Malnic, B., Godfrey, P. A. and Buck, L. B. (2004). The human olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7205-7205.

Malnic, B., Hirono, J., Sato, T. and Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell* **96**, 713-723.

Martin, G. R., Wilson, K., Wild, J. W., Parsons, S., Kubke, M. F. and Corfield, J. (2007). Kiwi forego vision in the guidance of their nocturnal activities. *PLoS ONE* **2**, e198.

Medway, L. (1959). Echo-location among collocalia. *Nature* **184**, 1352-1353.

Midford, P. E., Garland, T. J. and Maddison, W. P. (2005). PDAP package of Mesquite. Version 1.07. See http://mesquiteproject.org/pdap_mesquite/index.html.

Mighell, A. J., Smith, N. R., Robinson, P. A. and Markham, A. F. (2000). Vertebrate pseudogenes. *Febs Letters* **468**, 109-114.

Mombaerts, P. (1999a). Molecular biology of odorant receptors in vertebrates. *Annual Review of Neuroscience* **22**, 487-509.

Mombaerts, P. (1999b). Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**, 707-711.

Mombaerts, P. (2004). Genes and ligands for odorant, vomeronasal and taste receptors. *Nature Reviews Neuroscience* **5**, 263-278.

Moreno-Estrada, A., Casals, F., Ramirez-Soriano, A., Oliva, B., Calafell, F., Bertranpetit, J. and Bosch, E. (2008). Signatures of selection in the human olfactory receptor OR5I1 gene. *Molecular Biology and Evolution* **25**, 144-154.

Müller, J. C., Steiger, S., Fidler, A. E. and Kempenaers, B. (2008). Biogenic trace amine-associated receptors (TAARs) are encoded in avian genomes: Evidence and possible implications. *Journal of Heredity* **99**, 174-176.

Mundy, N. I. and Cook, S. (2003). Positive selection during the diversification of class I vomeronasal receptor-like (V1RL) genes, putative pheromone receptor genes, in human and primate evolution. *Molecular Biology and Evolution* **20**, 1805-1810.

N

Nef, S., Allaman, I., Fiumelli, H., De Castro, E. and Nef, P. (1996). Olfaction in birds: differential embryonic expression of nine putative odorant receptor genes in the avian olfactory system. *Mechanisms of Development* **55**, 65-77.

Nef, S. and Nef, P. (1997). Olfaction: Transient expression of a putative odorant receptor in the avian notochord. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4766-4771.

Nei, M. (1969). Gene duplication and nucleotide substitution in evolution. *Nature* **221**, 40-42.

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418-426.

Nei, M., Gu, X. and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7799-7806.

Nei, M. and Rooney, A. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* **39**, 121-151.

Nevitt, G. (2000). Olfactory foraging by Antarctic procellariiform seabirds: life at high Reynolds numbers. *Biological Bulletin* **198**, 245-253.

Nevitt, G. A., Losekoot, M. and Weimerskirch, H. (2008). Evidence for olfactory search in wandering albatross, Diomedea exulans. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4576-4581.

Ngai, J., Dowling, M. M., Buck, L., Axel, R. and Chess, A. (1993). The family of genes encoding odorant receptors in the channel catfish. *Cell* **72**, 657-666.

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J. et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PloS Biology* **3**, 976-985.

Nielsen, R. and Yang, Z. H. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929-936.

Niimura, Y. and Nei, M. (2003). Evolution of olfactory receptor genes in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12235-12240.

Niimura, Y. and Nei, M. (2005). Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6039-6044.

Niimura, Y. and Nei, M. (2006). Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *Journal of Human Genetics* **51**, 505-517.

Niimura, Y. and Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* **8**, e708.

Nozawa, M. and Nei, M. (2007). Evolutionary dynamics of olfactory receptor genes in Drosophila species. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7122-7127.

P

Papi, F. (1991). Olfactory navigation. In *Orientation in Birds*, (ed. P. Berthold), pp. 52-85. Basel: Birkhauser Verlag.

Parker, G. A. (1970). Sperm competition and its evolutionary consequences in insects. *Biological Reviews of the Cambridge Philosophical Society* **45**, 525-567.

Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gerard, C., Perret, J. et al. (1992). Expression of members of the putative olfactory receptor gene family in mammalian germ-cells. *Nature* **355**, 453-455.

Penn, D. J. (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. *Ethology* **108**, 1-21.

Petit, C., Hossaert-McKey, M., Perret, P., Blondel, J. and Lambrechts, M. M. (2002). Blue tits use selected plants and olfaction to maintain an aromatic environment for nestlings. *Ecology Letters* **5**, 585-589.

Petrie, M. and Kempenaers, B. (1998). Extra-pair paternity in birds: explaining variation between species and populations. *Trends in Ecology and Evolution* **13**, 52-58.

Pond, S. L. K. and Frost, S. D. W. (2005a). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533.

Pond, S. L. K. and Frost, S. D. W. (2005b). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.

Q

Quignon, P., Giraud, M., Rimbault, M., Lavigne, P., Tacher, S., Morin, E., Retout, E., Valin, A. S., Lindblad-Toh, K., Nicolas, J. et al. (2005). The dog and rat olfactory receptor repertoires. *Genome Biology* **6**, R83.

Quignon, P., Kirkness, E., Cadieu, E., Touleimat, N., Guyon, R., Renier, C., Hitte, C., Andre, C., Fraser, C. and Galibert, F. (2003). Comparison of the canine and human olfactory receptor gene repertoires. *Genome Biology* **4**, R80.

R

Robacker, D. C., Garcia, J. A. and Bartelt, R. J. (2000). Volatiles from duck feces attractive to Mexican fruit fly. *Journal of Chemical Ecology* **26**, 1849-1867.

Roper, T. J. (1999). Olfaction in birds. *Advances in the Study of Behaviour* **28**, 247-332.

Rouquier, S., Blancher, A. and Giorgi, D. (2000). The olfactory receptor gene repertoire in primates and mouse: Evidence for reduction of the functional fraction in primates. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 2870-2874.

S

Scheffler, K., Martin, D. P. and Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics* **22**, 2493-2499.

Sharon, D., Glusman, G., Pilpel, Y., Khen, M., Gruetzner, F., Haaf, T. and Lancet, D. (1999). Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics* **61**, 24-36.

Sibley, C. G. and Ahlquist, J. E. (1991). Phylogeny and classification of birds: A Study in Molecular Evolution. New Haven: Yale University Press.

Singer, M. S., Weisinger-Lewin, Y., Lancet, D. and Shepherd, G. M. (1996). Positive selection moments identify potential functional residues in human olfactory receptors. *Receptors and Channels* **4**, 141-147.

Spehr, M., Gisselmann, G., Poplawski, A., Riffell, J. A., Wetzel, C. H., Zimmer, R. K. and Hatt, H. (2003). Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science* **299**, 2054-2058.

Spehr, M., Schwane, K., Heilmann, S., Gisselmann, G., Hummel, T. and Hatt, H. (2004a). Dual capacity of a human olfactory receptor. *Current Biology* **14**, R832-R833.

Spehr, M., Schwane, K., Riffell, J. A., Barbour, J., Zimmer, R. K., Nauhaus, E. M. and Hatt, H. (2004b). Particulate adenylate cyclase plays a key role in human sperm olfactory receptor-mediated chemotaxis. *Journal of Biological Chemistry* **279**, 40194-40203.

Spehr, M., Schwane, K., Riffell, J. A., Zimmer, R. K. and Hatt, H. (2006). Odorant receptors and olfactory-like signaling mechanisms in mammalian sperm. *Molecular and Cellular Endocrinology* **250**, 128-136.

Steiger, S., Fidler A.E. and Kempenaers, B. (2008). Detection of olfactory receptor transcripts in bird testes. In press in *Journal of Heredity*.

Steiger, S., Fidler, A.D., Valcu, V. and Kempenaers, B. Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell? submitted to *Proceedings of the Royal Society of London Series B-Biological Sciences.*

Sun, H., Kondo, R., Shima, A., Naruse, K., Hori, H. and Chigusa, S. I. (1999). Evolutionary analysis of putative olfactory receptor genes of medaka fish, Oryzias latipes. *Gene* **231**, 137-145.

Suzuki, Y. and Gojobori, T. (1999). A method for detecting positive selection at single amino acid sites. *Molecular Biology and Evolution* **16**, 1315-1328.

T

Tacher, S., Quignon, P., Rimbault, M., Dreano, S., Andre, C. and Galibert, F. (2005). Olfactory receptor sequence polymorphism within and between breeds of dogs. *Journal of Heredity* **96**, 812-816.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F. and Higgins, D. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876-4882.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

V

Vanderhaeghen, P., Schurmans, S., Vassart, G. and Parmentier, M. (1993). Olfactory receptors are displayed on dog mature sperm cells. *J. Cell Biol.* **123**, 1441-1452.

Vanderhaeghen, P., Schurmans, S., Vassart, G. and Parmentier, M. (1997). Specific repertoire of olfactory receptor genes in the male germ cells of several mammalian species. *Genomics* **39**, 239-246.

Venables, W. N. and Ripley, D. (2002). Modern applied statistics with S: Springer Verlag.

W

Wallraff, H. G. (2004). Avian olfactory navigation: its empirical foundation and conceptual state. *Animal Behaviour* **67**, 189-204.

Warham, J. (1996). The behaviour, population biology and physiology of the petrels. London: Academic Press.

Warren, W. C. Hillier, L. W. Graves, J. A. M. Birney, E. Ponting, C. P. Grutzner, F. Belov, K. Miller, W. Clarke, L. Chinwalla, A. T. et al. (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175-184.

Wenzel, B. M. (1968). Olfactory prowess of kiwi. *Nature* **220**, 1133-1134.

Wilson, K.-J. (2004). Flight of the Huia: Ecology and conservation of New Zealand's frogs, reptiles, birds and mammals. Christchurch: Canterbury University Press.

Y

Yang, Z. and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* **15**, 496-503.

Yang, Z. H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**, 555-556.

Yang, Z. H. and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**, 32-43.

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L. and Trask, B. J. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Molecular Genetics* **11**, 535-546.

Young, J. M. and Trask, B. J. (2002). The sense of smell: genomics of vertebrate odorant receptors. *Human Molecular Genetics* **11**, 1153-1160.

Z

Zhang, X., Rogers, M., Tian, H., Zhang, X., Zou, D.-J., Liu, J., Ma, M., Shepherd, G. M. and Firestein, S. J. (2004a). High-throughput microarray detection of olfactory receptor gene expression in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14168-14173.

Zhang, X. H., De la Cruz, O., Pinto, J. M., Nicolae, D., Firestein, S. and Gilad, Y. (2007). Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biology* **8**, R86.

Zhang, X. M. and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nature Neuroscience* **5**, 124-133.

Zhang, X. M., Rodriguez, I., Mombaerts, P. and Firestein, S. (2004b). Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* **83**, 802-811.

SUMMARY

The sense of smell enables animals to e.g. locate food, to navigate or to avoid predators. In vertebrates, odorants are detected by olfactory receptors (ORs) that are expressed in the olfactory epithelium on olfactory sensory neurons. In this study, I investigated OR gene repertoires in both closely and distantly related bird species.

Using polymerase chain reaction (PCR) with degenerate primers designed to amplify OR genes, I showed that the majority of OR genes sequenced (~ 84%) were potentially functional in nine bird species from seven different orders. A nonparametric statistical technique was used to estimate the total number of OR genes in avian genomes. The total number of OR genes was surprisingly large and varied up to six fold between species (range: 106 – 667 OR genes). The total number of OR genes but not the proportion of potentially functional OR genes was positively correlated with the relative size of the olfactory bulb, which is considered an anatomical correlate of olfactory capability. A Southern Blot approach in combination with a PCR based approach revealed that two nocturnal bird species that heavily rely on olfactory cues, have evolved a larger OR gene repertoire than their diurnal, closest living relatives. Thus, it is likely that ecological niche adaptations (e.g. adaptations related to daily activity patterns) have shaped avian OR gene repertoires.

Phylogenetic trees derived from predicted OR protein sequences revealed that a large, expanded OR gene clade, termed group-γ-c, is present in all bird genomes examined in this study. This clade seems to be a shared characteristic of all bird genomes. Further, I showed that positive selection has driven the molecular evolution of avian group-γ-c OR genes. Positively selected sites encoded residues within transmembrane regions that most likely interact with odour molecules and thus might influence OR receptor functioning.

Interestingly, OR gene transcripts have been detected in testis and sperm of both mammals and fish, suggesting that OR genes are also involved in sperm-egg communication. Using reverse transcription (RT)-PCR with degenerate primers specific for OR genes, and subsequent cloning, I showed that several OR gene transcripts are present in chicken (*Gallus gallus domesticus*) testes and that they belong to the class-γ OR gene clade.

Finally, a database search in the red jungle fowl (*Gallus gallus*) genome revealed that trace amine-associated receptors (TAARs) - a second family of chemosensory receptors primarily expressed in the olfactory epithelium that detect amine-based odour cues - are also encoded in avian genomes.

The findings in this thesis contribute to our understanding of the evolution of avian OR genes. The estimated OR gene repertoire sizes, and the proportion of presumably functional OR genes, strongly suggest that avian olfactory ability is well developed and much more important than previously thought.

ACKNOWLEDGEMENTS

ADDRESSES OF CO-AUTHORS

Dr. Andrew E. Fidler

Cawthron Institute
Private Bag 2
Nelson, New Zealand
email: Andrew.Fidler@cawthron.org.nz

Prof. Dr. Bart Kempenaers

Max Planck Institute for Ornithology
Department of Behavioural Ecology &
Evolutionary Genetics
Eberhard-Gwinner-Strasse
82319 Seewiesen, Germany
email: b.kempenaers@orn.mpg.de

Dr. Jakob C. Mueller

Max Planck Institute for Ornithology
Department of Behavioural Ecology &
Evolutionary Genetics
Eberhard-Gwinner-Strasse
82319 Seewiesen, Germany
email: mueller@orn.mpg.de

Dr. Mihai Valcu

Max Planck Institute for Ornithology
Department of Behavioural Ecology &
Evolutionary Genetics
Eberhard-Gwinner-Strasse
82319 Seewiesen, Germany
email: valcu@orn.mpg.de

CURRICULUM VITAE

NAME   Silke Stefanie Steiger

PRESENT ADDRESS   Max Planck Institute for Ornithology
                      Department of Behavioural Ecology &
                      Evolutionary Genetics
                      Eberhard-Gwinner-Strasse
                      82319 Seewiesen, Germany
                      Tel: ++49 (0)8157 932 416
                      Fax: ++49 (0)8157 932 400
                      email: steiger@orn.mpg.de

PRIVATE ADDRESS   An den Meisterwiesen 2a
                      82229 Seefeld
                      Tel: ++49 (0)8152 395 787

DATE OF BIRTH   01.03.1978

PLACE OF BIRTH   Munich, Germany

SECONDARY EDUCATION

1990 - 1994   Therese-Giese Realschule, Unterschleißheim
1994 - 1998   Adolf-Weber Gymnasium, Munich

TERTIARY EDUCATION

1999 - 2004   Studies in biology
                      Ludwig-Maximilians-University
                      Munich, Germany
                      (Zoology, Genetics, Ecology, Immunology)

2004   Max Planck Institute for Ornithology
                      Diploma thesis
                      'Warum tropische Vögel länger leben: Freilandenergetik
                      eines Ameisenvogels im neotropischen Regenwald'
                      Supervisor: Prof. Dr. Martin Wikelski
                      (Princeton University, USA; present address: Max
                      Planck Institute for Ornithology, 78315 Radolfzell)

since Mar 2005   Max Planck Institute for Ornithology
                      Doctoral thesis
                      'Evolution of avian olfaction'
                      Supervisors: Prof. Dr. Bart Kempenaers (Max Planck
                      Institute for Ornithology), Dr. Andrew Fidler (Cawthron
                      Institute, New Zealand)

SCHOLARSHIPS

Mar – Apr 2003	Hochschulfonds München
Aug – Oct 2003	Deutscher Akademischer Austauschdienst (DAAD)
Jan 2008	Deutscher Akademischer Austauschdienst (DAAD)

WORKING EXPERIENCE

Feb – Oct 1999   Practical assistant
Mar – Apr 2000   Bavarian Nordic Research Institute
Mar – Apr 2001   Martinsried, Germany
                     Supervisors: Dr. Robert Saller, Dr. Sonja Leyrer,
                     Frank Heinemann

Aug – Oct 2000   Practical assistant
                     University of Queensland
                     Centre for Studies in Drug Disposition
                     Brisbane, Australia
                     Supervisor: Dr. Andrew King

Dec 2000 – Jan 2001 Field assistant
                     Galapagos Islands, Ecuador
                     Supervisor: Prof. Dr. Martin Wikelski

Feb – Mar 2002   Practical assistant
                     Max Planck Institute for Ornithology
                     Seewiesen, Germany
                     Supervisor: Prof. Dr. Herbert Biebach

Jun – Jul 2004   Research assistant
Jun 2005   Barrow, Alaska, USA
Jun 2006   Supervisor: Prof. Dr. Bart Kempenaers
Jun 2007

PUBLICATIONS

Steiger, S., Fidler A.E. and Kempenaers, B.** (2008). Detection of olfactory receptor transcripts in bird testes. In press in *Journal of Heredity*.

Mueller, J.C., Steiger, S., Fidler A.E. and Kempenaers, B. (2008). Biogenic trace amine-associated receptors (TAARs) are encoded in avian genomes: evidence and possible implications. *Journal of Heredity*, **99**, 174-176.

Steiger, S., Goymann, W. and Kempenaers, B. (2006). Plasma steroid hormones in two Arctic-breeding shorebirds: monogamy versus polygyny. *General and Comparative Endocrinology*, **147**, 133-140.

Wikelski, M., <u>Steiger, S.</u>, Gall, B. and Nelson, K. (2005). Sex, drugs and mating role: testosterone-induced phenotype-switching in Galapagos marine iguanas. *Behavioral Ecology*, **16 (1)**, 260-268.

Mettke-Hofmann, C., Ebert, C., Schmidt, T., <u>Steiger, S.</u> and Stieb, S. (2005). Personality traits in resident and migratory warbler species. *Behaviour*, **142**, 1363-1381.

Ehrenwörtliche Versicherung

Ich versichere hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den ....   ……………...

(Unterschrift)

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

München, den.............................   ....

(Unterschrift)