GBE

Genome Biology and Evolution

SMBE

Contrasting Patterns of Evolutionary Diversification in the Olfactory Repertoires of Reptile and Bird Genomes

Michael W. Vandewege[1], Sarah F. Mangum[2], Toni Gabaldón[3,4,5], Todd A. Castoe[6], David A. Ray[2], and Federico G. Hoffmann[1,7,*]

[1]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University

[2]Department of Biological Sciences, Texas Tech University

[3]Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain

[4]Universitat Pompeu Fabra (UPF), Barcelona, Spain

[5]Institució Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Spain

[6]Department of Biology, University of Texas at Arlington

[7]Institute for Genomics, Biocomputing and Biochemistry, Mississippi State University

*Corresponding author: E-mail: federico.g.hoffmann@gmail.com.

Accepted: January 28, 2016

Abstract

Olfactory receptors (ORs) are membrane proteins that mediate the detection of odorants in the environment, and are the largest vertebrate gene family. Comparative studies of mammalian genomes indicate that OR repertoires vary widely, even between closely related lineages, as a consequence of frequent OR gains and losses. Several studies also suggest that mammalian OR repertoires are influenced by life history traits. Sauropsida is a diverse group of vertebrates group that is the sister group to mammals, and includes birds, testudines, squamates, and crocodilians, and represents a natural system to explore predictions derived from mammalian studies. In this study, we analyzed olfactory receptor (OR) repertoire variation among several representative species and found that the number of intact OR genes in sauropsid genomes analyzed ranged over an order of magnitude, from 108 in the green anole to over 1,000 in turtles. Our results suggest that different sauropsid lineages have highly divergent OR repertoire composition that derive from lineage-specific combinations of gene expansions, losses, and retentions of ancestral OR genes. These differences also suggest that varying degrees of adaption related to life history have shaped the unique OR repertoires observed across sauropsid lineages.

Key words: ORs, Sauropsida, gene family evolution, gene duplication.

Introduction

In vertebrates, the ability to detect odors is mediated by ORs, a type of transmembrane G protein-coupled receptor (GPCR) that mediates interactions between the cell and its surroundings. Structurally, GPCRs have seven α-helical transmembrane domains bound to a G-protein, and the binding of extracellular ligands triggers conformational changes that, in turn, lead to intracellular signaling cascades (Fredriksson et al. 2003). Vertebrate ORs belong to the rhodopsin-like group of GPCRs, which includes receptors that mediate the detection of hormones, neurotransmitters, and photons (Fredriksson et al. 2003). Vertebrate ORs are primarily expressed in the olfactory epithelium of the nasal cavity, where they bind odorants, and transmit the resulting nerve impulse to the brain (Buck and Axel 1991; Mombaerts 1999). The OR repertoires of amniote vertebrates are dominated by two major groups of ORs, Class I ORs, which appear to have a higher affinity for hydrophilic ligands, and Class II ORs, which generally bind hydrophobic ligands (Saito et al. 2009).

Genomic surveys have revealed that ORs represent the largest vertebrate gene family (Zhang and Firestein 2002), and indicate that the numbers and diversity of ORs vary widely among vertebrates, even between closely related taxa (Niimura and Nei 2005b; Nei et al. 2008). There is debate regarding the relative influence of different evolutionary forces in shaping OR repertoires. Nei et al. (2008) suggests that OR evolution is largely a neutral process, whereas multiple comparative studies report that similarities among OR repertoires reflect shared ecology and anatomy rather than phylogenetic relatedness (Hayden et al. 2010, 2014; Garrett and

Steiper 2014; Khan et al. 2015). Consistent with the prominent roles of ecology and anatomy, the size of the OR repertoire has been previously related to reliance on olfaction. There are approximately 800 OR genes in the human genome, half of which appear to be pseudogenes, whereas there are more than 1,000 intact OR genes in the mouse genome and approximately 2,000 intact ORs in the elephant (Glusman et al. 2001; Zhang and Firestein 2002; Niimura et al. 2014). Further, although clear links between particular ORs and specific chemical ligands are largely missing, multiple studies have linked features of the OR repertoires to ecological adaptation and lineage-specific specialization (Steiger et al. 2009; Hayden et al. 2010, 2014; Garrett and Steiper 2014; Niimura et al. 2014; Khan et al. 2015).

Most of the comparative studies of the OR repertoires of tetrapods have focused on mammals because of the greater availability of mammalian genome drafts (Zhang and Firestein 2002; Niimura and Nei 2003, 2007; Hayden et al. 2010, 2014; Matsui et al. 2010; Niimura et al. 2014), with a recent study comparing bird OR repertoires as a notable exception (Khan et al. 2015). Sauropsids are the sister group of mammals, and include Rhynchocephalia (tuatara), Squamates (snakes and lizards), Testudines (turtles and tortoises), and Archosaurs (crocodilians, dinosaurs, and birds), and with the exception of birds, have been largely absent from OR studies. Multiple genomes from representatives from this group have been released recently (Castoe et al. 2013; Wan et al. 2013; Wang et al. 2013; Green et al. 2014) and offer an opportunity to explore the evolution of OR repertoires in amniote vertebrate lineages other than mammals. Therefore, the primary goal of our study was to investigate patterns of diversification of sauropsid OR repertoires using these recently released genomes.

Prior studies based on the genomic analyses of the green anole, chicken, and zebra finch suggest that squamates have smaller OR repertoires than most mammals (Steiger et al. 2009) and that gene loss played a prominent role in the evolution of avian OR repertoires (Khan et al. 2015). Similarly, the OR repertoires of birds appear to be small relative to most other amniotes yet include an expansion of OR subfamily 14 (Lagerström et al. 2006; Steiger et al. 2008, 2009; Khan et al. 2015). The phyletic extent of this expansion has not, however, been fully resolved. Further, it is not known whether snakes, which rely heavily on their sense of smell and chemoreception abilities (Cooper 1991; Stone and Holtzman 1996; Shine and Mason 2001; LeMaster and Mason 2002; Clark 2007), do indeed have a reduced OR repertoire like that observed in the green anole. Similarly, the OR repertoires in crocodiles and turtles, which invaded semiaquatic niches independently, have yet to be thoroughly analyzed and compared. Because Class I ORs are thought to be primarily involved in detecting aquatic-borne odorants and are particularly abundant in turtles (Wang et al. 2013), we were interested in evaluating whether semiaquatic crocodilians may have also experienced an expansion of Class I ORs. To address these questions, we analyzed patterns of OR gene gain and loss from a sample of sequenced sauropsid genomes. Our results indicate that different sauropsid lineages have diverse OR repertories that range from few to several hundred genes derived from lineage-specific combinations of expansions, losses, and differential retention of ancestral genes.

Materials and Methods

Data Sources

We queried the genomes for putative ORs from the following representative sauropsid species: green anole (*Anolis carolinensis*), Burmese python (*Python morulus*), Chinese softshell turtle (*Pelodiscus sinensis*), painted turtle (*Chrysemys picta*), American alligator (*Alligator mississippiensis*), Indian gharial (*Gavialis gangeticus*), saltwater crocodile (*Crocodylus porosus*), chicken (*Gallus gallus*), and zebra finch (*Taeniopygia guttata*). We included duckbilled platypus (*Ornithorhynchus anatinus*) as an outgroup. Additional genome details are provided in supplementary table S1, Supplementary Material online. Although many of these genome drafts have been previously surveyed for ORs, we reannotated these genomes to benchmark the accuracy of our OR prediction approach, and to provide a consistent basis for the annotation of ORs across genomes for comparative analyses. Further, in many of these cases only OR numbers were reported, therefore we sought to provide more detail regarding subfamily designations and comparative evolutionary histories among OR repertoires which has yet to be conducted.

OR Prediction

To identify putative ORs, we implemented a bioinformatic pipeline similar to the one described in Niimura and Nei (2007). Briefly, we conducted TBLASTN searches of the specified genomes excluding hits with an e-value greater than 1e-10. These searches were conducted using as queries a set of known ORs from the green anole, African clawed frog (*Xenopus tropicalis*), chicken, and zebra fish (*Danio rerio*) from Niimura (2009), and human ORs from Niimura and Nei (2003). Hits shorter than 150 bp were discarded. We extracted the best BLAST hits identified by the smallest e-value from nonoverlapping regions, plus 999 bp in the upstream and downstream flanking sequences, using modules in BEDTOOLS (Quinlan and Hall 2010) and custom Python scripts. Putative OR genes were considered intact if there was an uninterrupted open reading frame (ORF) with no gaps $\geq$5 amino acids in the seven transmembrane domains or conserved regions, and an appropriate stop codon. Newly discovered intact ORs were added to the amino acid query and the TBLASTN search was conducted a second time to discover potentially undetected pseudogenes and truncated genes using a cutoff of 1e-20. The best hits, plus 99 bp upstream and downstream, were extracted. ORs were

considered pseudogenes if the longest ORF was shorter than 250 amino acids, there were gaps of five or more amino acids in the transmembrane domains or conserved regions, frameshift mutations, or premature stop codons. OR sequences located at the end of a scaffold or interrupted by scaffold gaps, but otherwise apparently intact, were considered truncated. Truncated ORs were validated by alignment to functional genes using MAFFT 7.127 (Katoh and Toh 2008) and visually inspected for premature stop codons and gaps within conserved regions. Predicted OR amino acid sequences were mapped back to their corresponding genome to annotate their precise coordinates and orientation.

Class I and II ORs diverged and diversified early in tetrapod evolution (Niimura 2009). Mammalian OR genes have been historically classified into 18 subfamilies, 4 Class I subfamilies (51, 52, 55, 56) and 14 Class II subfamilies identified from the human genome (Glusman et al. 2000). However, Hayden et al. (2010) determined that several of the previously classified Class II subfamilies were not monophyletic among all mammals and subsequently defined new groups by identifying monophyletic lineages of ORs (1/3/7, 2/13, 4, 5/8/9, 11, 6, 10, 12, 14). We used BLASTP to group intact ORs into putative subfamilies based on human ORs and the classifications of Hayden et al. (2010). We then verified and corrected the putative BLASTP-based assignments based on the inferred phylogenetic tree of the full OR dataset (see below). Intact OR amino acid sequences are available as part of the supplementary material, Supplementary Material online. We assigned pseudogenes to OR subfamilies in the following manner. We created a database of all of the annotated amino acid sequences and used BLASTX to query the pseudogene nucleotide sequences against the protein database. We used a cutoff of 1e-10 and allowed ten target sequences per query sequence. The subfamily annotation that was most frequent among the ten hits was assigned to the pseudogene.

Analyses

After annotation, we used CAFÉ (De Bie et al. 2006) to reconstruct the OR repertoires from the number of intact Class I and Class II genes to identify ancestral OR gene copy number states given the gene gain and loss in each lineage. The CAFÉ method assumes equal probability of birth (duplication) and death (deletion/pseudogenization). Divergence times for each node in the CAFÉ analyses were taken from TimeTree (Hedges et al. 2006).

We estimated the evolutionary relationships of OR sequences based on amino acid alignments. In all cases, we aligned the amino acid sequences of intact ORs using EINSI parameters in MAFFT 7.127. We created a full alignment of all intact ORs and also separate alignments of OR sequences for the birds, crocodilians, turtles, and squamates, and estimated phylogenetic relationships using Fastree2 (Price et al. 2010), which is specifically designed to calculate "approximately maximum-likelihood" phylogenetic trees on extremely large alignments such as those generated from aligning thousands of ORs here. Nodal support was estimated from 1,000 bootstrap replicates. The resulting tree was used to infer and date gene duplication events based on a phylogeny-aware algorithm (Huerta-Cepas and Gabaldón 2011) as implemented in ETE v2 (Huerta-Cepas et al. 2010). This method is complementary to CAFÉ, which does not consider the topology of the gene tree.

In most vertebrates studied to date, OR genes are spatially clustered (Giglio et al. 2001; Niimura and Nei 2005a). Thus, it was of interest to investigate how ORs were organized and distributed across various sauropsid genomes. To do so, we analyzed spatial clustering patterns of genetically linked OR genes using BEDTOOLS to locate genomic clusters of ORs in each genome in our analysis, even though establishing the exact boundaries of OR clusters was difficult for most genome drafts. OR clusters can be several Mb long yet many of the unmapped scaffolds containing ORs were shorter than 1 Mb due to the overall shorter scaffold sizes of some genome assemblies we analyzed. Due to this limitation, we defined clusters as three or more OR genes that are separated by less than 100 kb of one another. Clusters that were within 10 kb of a scaffold end were considered incomplete.

Results and Discussion

We first compared results from our bioinformatic pipeline on updated drafts of the green anole, zebra finch, and chicken with the original reports. We found that gene counts were similar between anoCar1 and anoCar2, that galGal4 had more ORs than galGal3, and that our counts were very similar to those in the zebra finch and softshell turtle reported in Wang et al. (2013) (supplementary table S2, Supplementary Material online). Our annotation of the python genome yielded more ORs than previous estimates (Dehara et al. 2012; Castoe et al. 2013). Overall, these comparisons suggest that our pipeline generates results that are generally comparable to those from previous studies, and in some cases more inclusive. Thus, we infer our characterization of the OR repertoires of painted turtle, python, gharial, American alligator, and saltwater crocodile represent robust estimates of the diversity and size of the OR gene family in these genomes.

OR Repertoires Vary among Major Sauropsid Groups

Quantitative comparisons of ORs across genomes indicate that sauropsids evolved extensive variation in the size of the OR repertoires, as the number of intact genes in the genomes analyzed ranged over an order of magnitude, from 108 in the green anole to 1,180 in the Chinese softshell turtle. Similarly, the number of pseudogenes ranged from 33 in the green anole to 538 in the American alligator (table 1) and the number of truncated but putatively coding genes ranged from one in the green anole to 598 in the python.

Table 1
Summary of OR Gene Annotations from Each Genome

Genome	Intact (I)	Pseudogenes (P)	Truncated (T)	Total (I+P+T)	%Truncated (T/I+T)	%Intact (I/I+P)
Platypus	270	351	35	656	11	44
Green Anole	108	33	1	142	0.9	77
Python	481	319	598	1398	55	60
Softshell Turtle	1180	533	40	1753	3	69
Painted Turtle	842	942	279	2063	24	47
Crocodile	592	331	66	989	10	64
Gharial	597	389	153	1139	18	61
Alligator	465	538	74	1077	14	46
Zebra Finch	190	306	45	541	19	38
Chicken	266	173	83	522	24	61

The abundance of these truncated genes did not appear to be related to the overall contiguity of genome assembly, since the crocodile and gharial genomes had shorter scaffold N50s yet fewer truncated genes (table 1; supplementary table S1, Supplementary Material online).

Intriguingly, the two squamates in this study, python and green anole, diverged approximately 160 Ma (Evans 2003; Castoe et al. 2009) and exhibit the largest difference in the number of ORs between species within a major sauropsid lineage, 481 in the python to 108 in the anole (table 1). This difference is probably higher, as the number of truncated genes in the python genome (table 1 and fig. 1*A*) suggests the number of intact genes in the python genome is likely higher than our current estimate. Despite these numerical differences, both species have repertoires dominated by Class II ORs (fig. 1*A*) with similar subfamily proportions in the two species (fig. 1*B*).

The two testudines in our study, Chinese softshell and painted turtle, diverged approximately 170 Ma (Pyron 2010). The turtle genomes contained the largest numbers of intact ORs among sauropsids, and included several hundred Class I genes (fig. 1*A*), primarily from subfamily 52 (fig. 1*B*). This class of ORs is thought to mediate detection of waterborne odorants (Saito et al. 2009). Compared to the Chinese softshell turtle, the painted turtle genome contained a higher fraction of truncated ORs (24% vs. 3% in the softshell turtle, fig. 1*A*) and pseudogenes (~50% vs. ~30% in the softshell turtle).

The two extant groups of Archosaurs, birds and crocodilians, show marked differences in their OR repertoires. Chicken and zebra finch had the second smallest number of ORs, with 200 and 250 ORs, respectively, almost all of which belonged to subfamily 14 (fig. 1*B*). In contrast, crocodilian genomes encode more than twice the number of intact ORs, between 465 and 597 (table 1), derived from multiple subfamilies (fig. 1*B*). It is notable that although the three crocodilian species diverged approximately 90 Ma (Roos et al. 2007), they have similar OR repertoires in terms of gene numbers (fig. 1*A*) and subfamily composition (fig. 1*B*), further illustrating suggestions that crocodilian genomes have remained remarkably static and conserved over many millions of years (Green et al. 2014).

OR Pseudogenization

If there has been no gene gain and pseudogenes are retained in the genome, there should be a negative correlation between the number of intact ORs and number of pseudogenes. To test this prediction, we explored the number of pseudogenes and their distribution across OR subfamilies. We calculated the proportion of pseudogenes by dividing the number of pseudogenes by the total number of genes, excluding truncated genes because they cannot be classified confidently. Our analyses indicate that the overall proportion of pseudogenes was not correlated with the number of intact genes (fig. 2*A*). However, we did find a significant positive correlation between the proportion of pseudogenes per subfamily and the proportion of intact genes per subfamily ($r^2 = 0.87$, $P < 0.0001$, fig. 2*B*). These two observations together suggest that the pseudogenes present reflect the composition of the OR repertoire, but that the current abundance of pseudogenes is not determined by the abundance of intact genes. The fraction of pseudogenes can change when genes or pseudogenes are deleted from the genome (Niimura et al. 2014) and although several groups have suggested that the proportion of pseudogenes relative to the total number of genes is related to olfactory ability (Kishida et al. 2007; Hayden et al. 2010; Kishida and Hikida 2010), our results are not consistent with this. In agreement with Niimura et al. (2014), our analyses suggest that the fraction of pseudogenes is a poor indicator of olfactory ability.

Genomic Organization of OR Genes

In most mammalian genomes, OR genes are arranged in gene clusters composed of closely related genes and orthologous clusters are often shared among relatively distantly related species, such as the human and the mouse (Niimura and Nei 2005a). Similarly in sauropsids, the proportion of ORs in

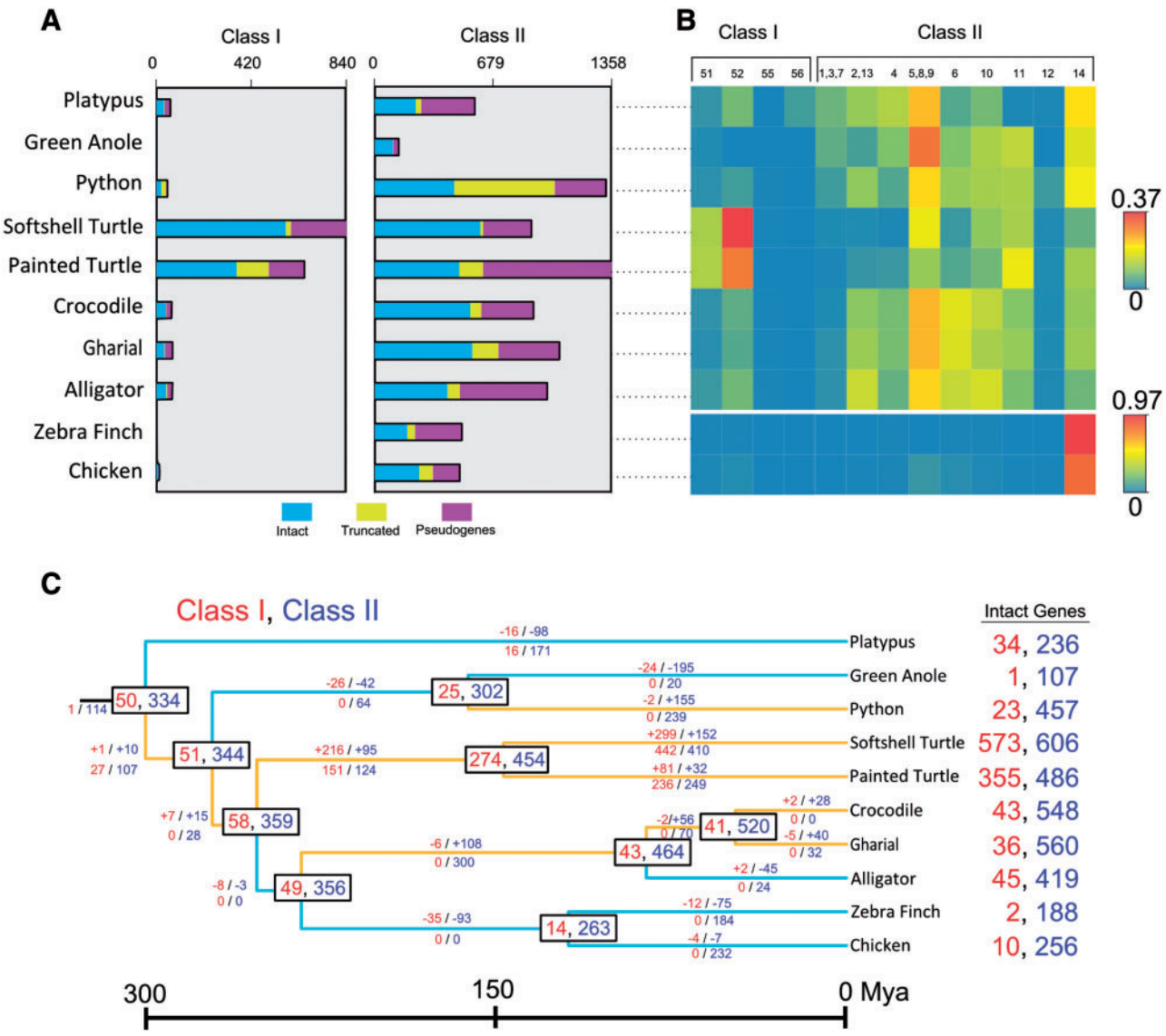

Fig. 1.—(*A*) The comprehensive collection of annotated Class I and Class II genes in each taxon. (*B*) Heat map based on the proportion of intact ORs that belong to OR subfamilies. Avian and nonavian groups were presented on two different scales because more than 85% of avian ORs are in subfamily 14, whereas the highest percentage is 36% in subfamily 52 of turtles. (*C*) Historic Class I and Class II gene numbers in the ancestral nodes and gain/loss along each branch of taxa (CAFÉ analysis, above branches), and the inferred number of past duplication events per OR Class and lineage, based on the gene phylogeny and a species-overlap duplication detection and dating algorithm (Huerta-Cepas and Gabaldón (2011), below branches). Light blue branches are those with an average gene loss per Class and orange branches are those with an average gene gain.

clusters ranged from 42% to 90%, in the zebra finch and softshell turtle, respectively, and the number of OR gene clusters per genome ranged from 5 to 139 across species (table 2). Some of these clusters were composed of a single subfamily, but most clusters (such as the largest cluster in the painted turtle) contained multiple OR subfamilies (fig. 3*A*). As expected, genome drafts with lower scaffold N50s exhibited smaller clusters and greater abundances of incomplete clusters (table 2; supplementary table S1, Supplementary Material online). True OR cluster sizes are likely larger than our estimates (due to the fragmentary nature of assemblies), and ultimately the majority of ORs may be located in a small number of clusters, as in the green anole where almost 85% of ORs were found in only five clusters (table 2). For example, five scaffolds contained the majority of subfamily 51 ORs in the softshell turtle (fig. 3*B*), and the entire length of these five contigs is composed almost exclusively of these ORs (fig. 3*C*). The combined length of these five contigs is approximately 1.5 Mb. Because almost all of the subfamily 51 ORs are scattered among these five contigs and each of these contigs is

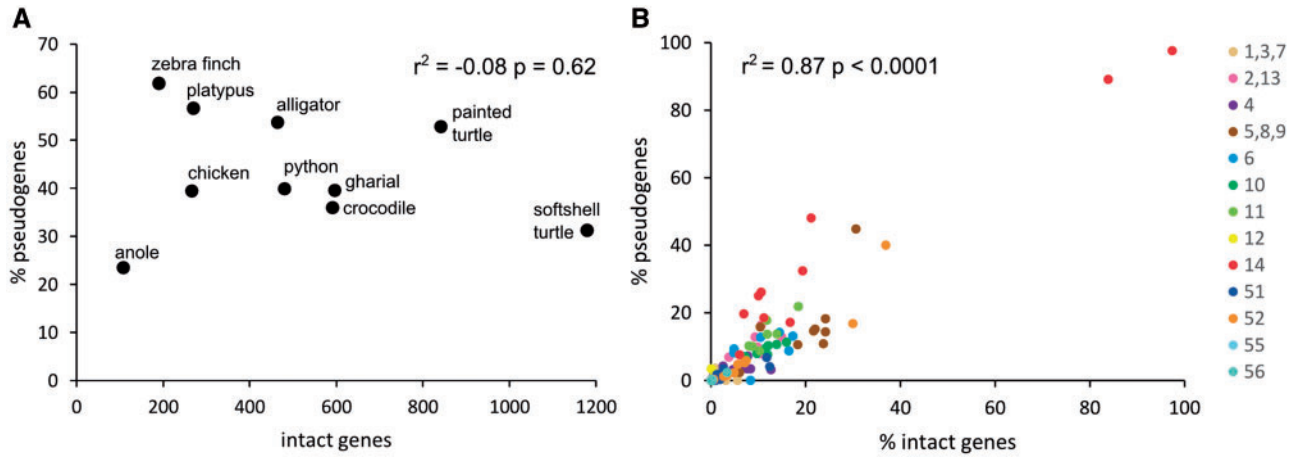

FIG. 2.—(*A*) The number of intact genes plotted against the percentage of pseudogenes as a proportion of the total number of intact genes and pseudogenes (p/p + i) within the same genome. A simple linear model was applied to the data and there was no significant correlation. (*B*) We plotted the percentage of pseudogenes against the percentage of intact genes for all subfamilies in all species and again applied a simple linear model to the data and found a strong linear relationship between the two metrics.

Table 2
Summary of OR Gene Clusters

Genome	Clusters	Complete	5′-Incomplete	3′-Incomplete	5′- and 3′-Incomplete	% Genes in Clusters
Platypus	**39**	**11**	**23**	**1**	**4**	**41**
Green Anole	**5**	**5**	**0**	**0**	**0**	**83**
Python	**130**	**16**	**84**	**30**	**0**	**58**
Softshell Turtle	**126**	**30**	**83**	**3**	**10**	**90**
Painted Turtle	**115**	**53**	**51**	**6**	**5**	**78**
Crocodile	**122**	**30**	**79**	**2**	**11**	**69**
Gharial	**139**	**48**	**67**	**12**	**12**	**58**
Alligator	**120**	**0**	**58**	**3**	**15**	**75**
Zebra Finch	**52**	**52**	**0**	**0**	**0**	**42**
Chicken	**27**	**7**	**18**	**0**	**2**	**57**

incomplete at the 5′ and 3′-end, we suspect these five contigs may represent a single contiguous cluster, containing the majority of the subfamily 51 ORs. The largest single human OR locus contains approximately 130 ORs and the largest mouse locus contains approximately 250 ORs (Niimura and Nei 2005a). By comparison, the largest clusters in our study were observed in the anole, platypus, painted turtle, and softshell turtle (assuming a single contiguous cluster) and contained 75, 93, 108, and 147 total ORs, respectively. These comparisons should be considered preliminary, as the actual cluster sizes will likely change as future more contiguous genome assemblies become available. Regardless of the exact numbers of clusters per genome, observed patterns of OR clustering across genomes suggest that tandem duplication is a primary source of novel ORs, as suggested by Niimura and Nei (2005a).

Evolution of OR Repertoires

We tracked gene gain and loss for Class I and Class II ORs on the species tree using maximum-likelihood estimates of OR repertoire size (fig. 1*C*). These analyses suggest that the relatively small OR repertoires of the chicken, the zebra finch, and the anole are the result of multiple gene losses, and that ancestors of both the birds and the green anole likely contained larger OR repertoires. Interestingly, the analysis of the anole and the two birds yield drastically different patterns of gene loss. The anole lost ORs from all subfamilies while retaining repertoire diversity, whereas the two birds analyzed lost almost all ORs belonging to all subfamilies except subfamily 14. Khan et al. (2015) previously demonstrated that birds generally have diverse OR repertoires. The notable exceptions were chicken, zebra finch, and the little egret, as more than 90% of the ORs in these genomes were made of subfamily 14

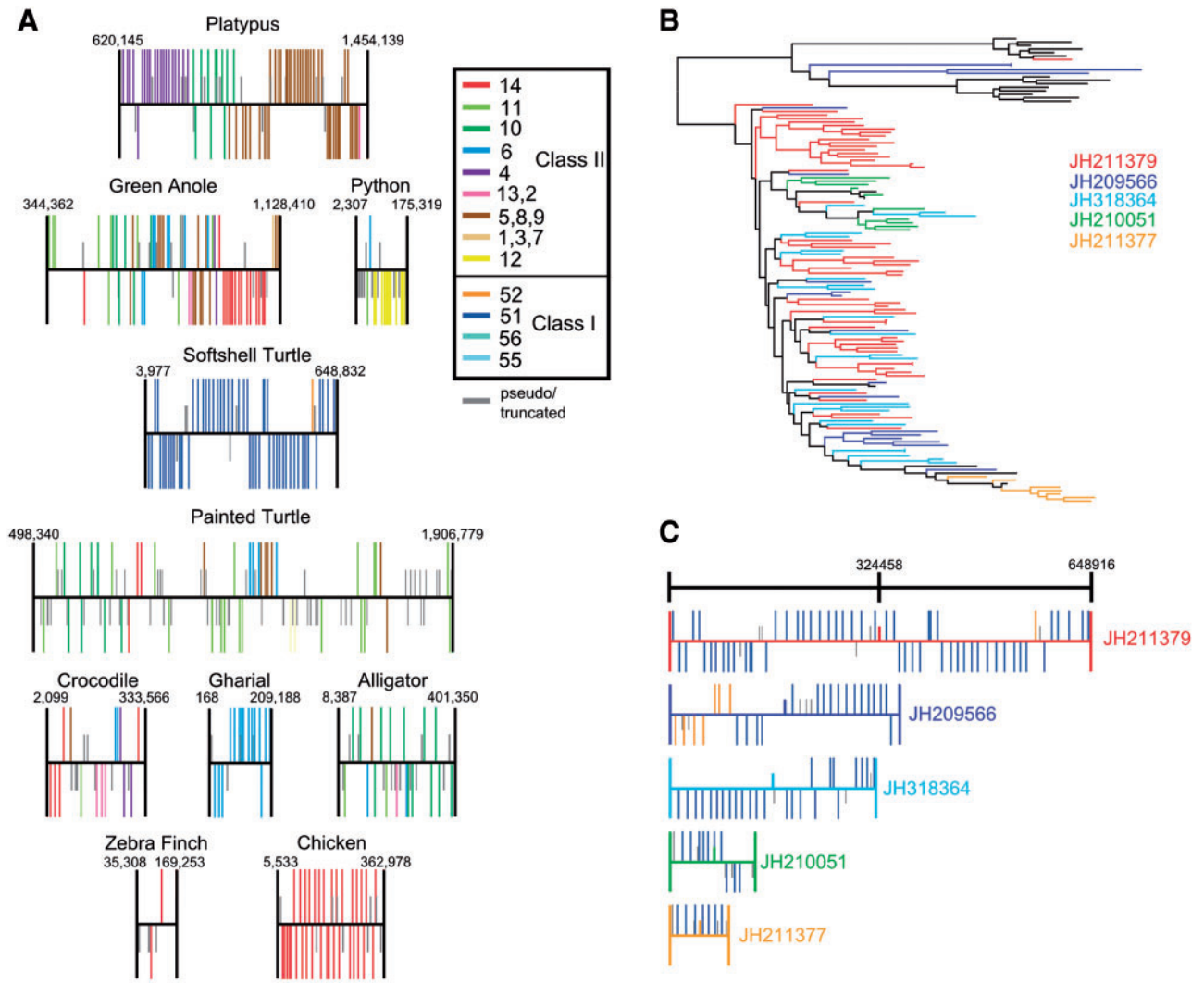

Fig. 3.—(*A*) The largest (most numerous) OR gene cluster from each genome draft. Each vertical bar represents a position of an OR. Bars above the horizontal line represent sense oriented genes and bars below the line represent antisense oriented genes in relation to the scaffold sequences. Each OR is colored according to the annotated subfamily. Cluster lengths are drawn to scale. (*B*) Neighbor joining tree of the subfamily 51 ORs in the softshell turtle; branches are colored according to the contig each OR was identified on. (*C*) The OR content of each contig presented in the *B* panel. Contig lengths are to scale, and the gene color scheme is congruent with the legend in panel *A*.

ORs suggesting that almost all OR subfamilies were lost independently in these three species.

Reconstructions of the OR repertoire in the common ancestor of sauropsids suggest it had 51 Class I and 344 Class II ORs. Ancestral nodes also had hundreds of Class II receptors, ranging from 263 to 520, and tens of Class I receptors, from 14 to 58, with the exception of the common ancestor of softshell and painted turtle, which had an estimated 274 Class I receptors. Turtles are notable because they are the only group analyzed to have gained Class I ORs at a greater rate than Class II ORs (fig. 1*C*). The crocodilian ancestor is estimated to have gained approximately 100 ORs since diverging from birds, and interestingly, the number of ORs in the three crocodilians has apparently remained remarkably similar to the number inferred for their ancestor.

To investigate patterns of OR gene gain and loss among subfamilies in more detail, we estimated phylogenetic relationships among all 4,991 intact OR genes (fig. 4). Class I and Class II ORs formed highly supported monophyletic clades (fig. 4*A*). Most mammalian-defined subfamilies generally formed monophyletic groups that included mammalian and sauropsid representatives (fig. 4*A*). Exceptions to this pattern include subfamily 5/8/9, which is paraphyletic in our analysis. The 5/8/9 subfamily is represented by three relatively distant clades (fig. 4*A*) that each includes representatives of all sauropsids (figs. 1*B* and 4*B*). Because of the difficulties in

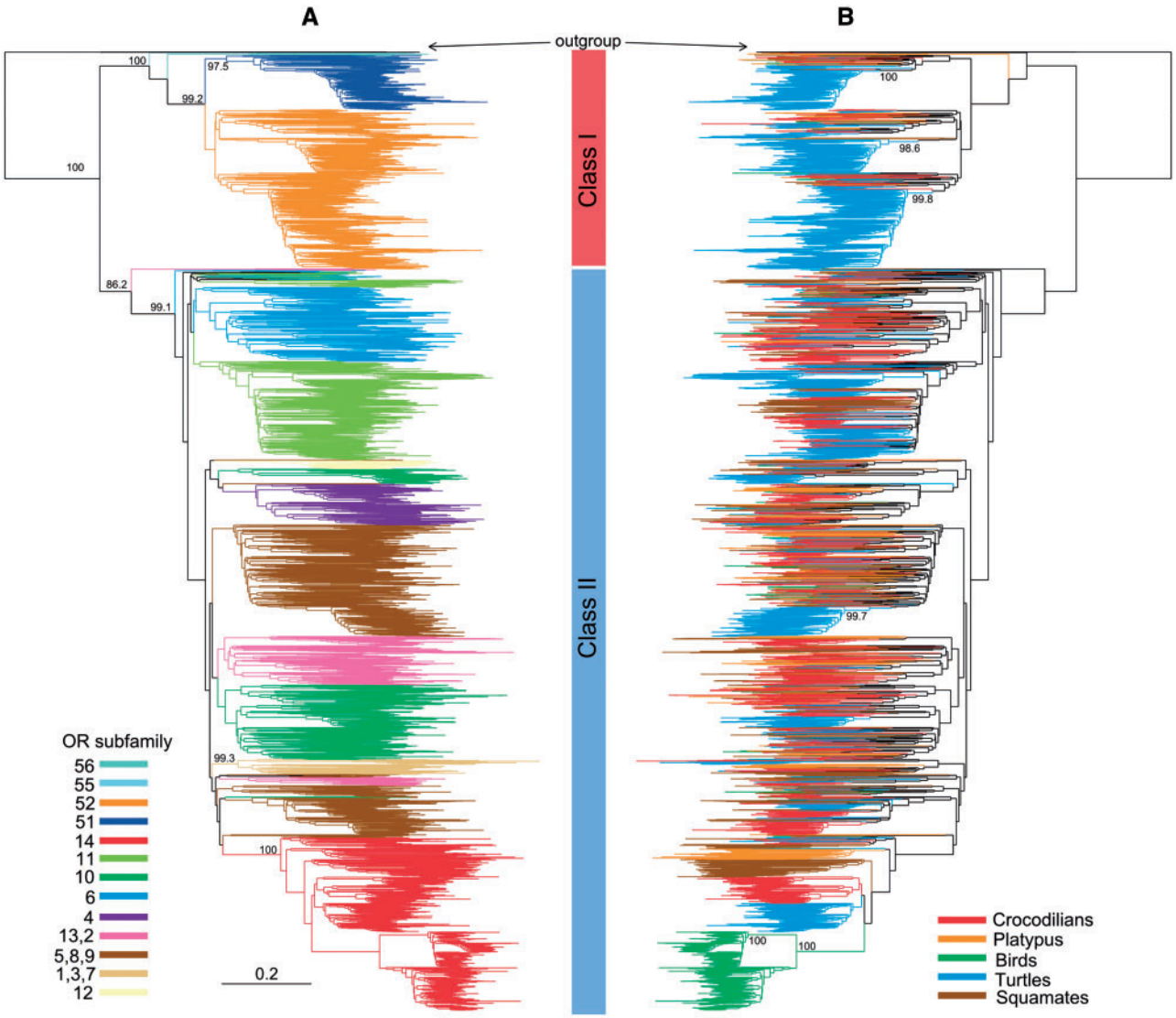

Fig. 4.—Phylogenetic tree estimate of the 4,991 intact ORs. (*A*) Branches colors are based on the annotated OR subfamily, nodal support is listed for the Classes and high supported subfamilies. (*B*) The same tree presented in A but branches are colored according to the major taxonomic classifications and nodal support is presented for high supported group-specific OR expansions.

resolving a tree with many more sequences than sites in the alignment, we restricted our primary focus to strongly supported monophyletic OR subfamilies, such as groups 51, 52, and 14 (fig. 4*A*).

In most cases, it was uncommon for ORs from a species or lineage to form a monophyletic group within a subfamily suggesting that these subfamilies had expanded prior to radiation of these species. Subfamily 14 was an exception, as almost all ORs in this subfamily formed species or lineage-specific clades (fig. 4*B*) suggesting that the same ancestral gene expanded independently multiple times in different lineages. Within this subfamily, chicken and zebra finch ORs are the most remarkable, as they are reciprocally monophyletic, stemming exclusively from species-specific expansions (fig. 5C; Khan et al. 2015), with a long branch leading to their common ancestor gene (fig. 4*B*).

When we constructed independent phylogenetic trees for each major sauropsid group, we were able to visualize the distinct patterns of gene gain and loss that produced current OR repertoires (fig. 5). Phylogenetic tree characteristics tended to be fundamentally different among the four groups. Turtles and crocodilian genomes are both notable for evolving slowly (Shaffer et al. 2013; Green et al. 2014), yet the OR repertoires of turtles show extensive evolutionary dynamics with multiple species-specific expansions (fig. 5*B*) while crocodilian OR repertoires have apparently experienced little change in gene

Genome Biology and Evolution

SMBE

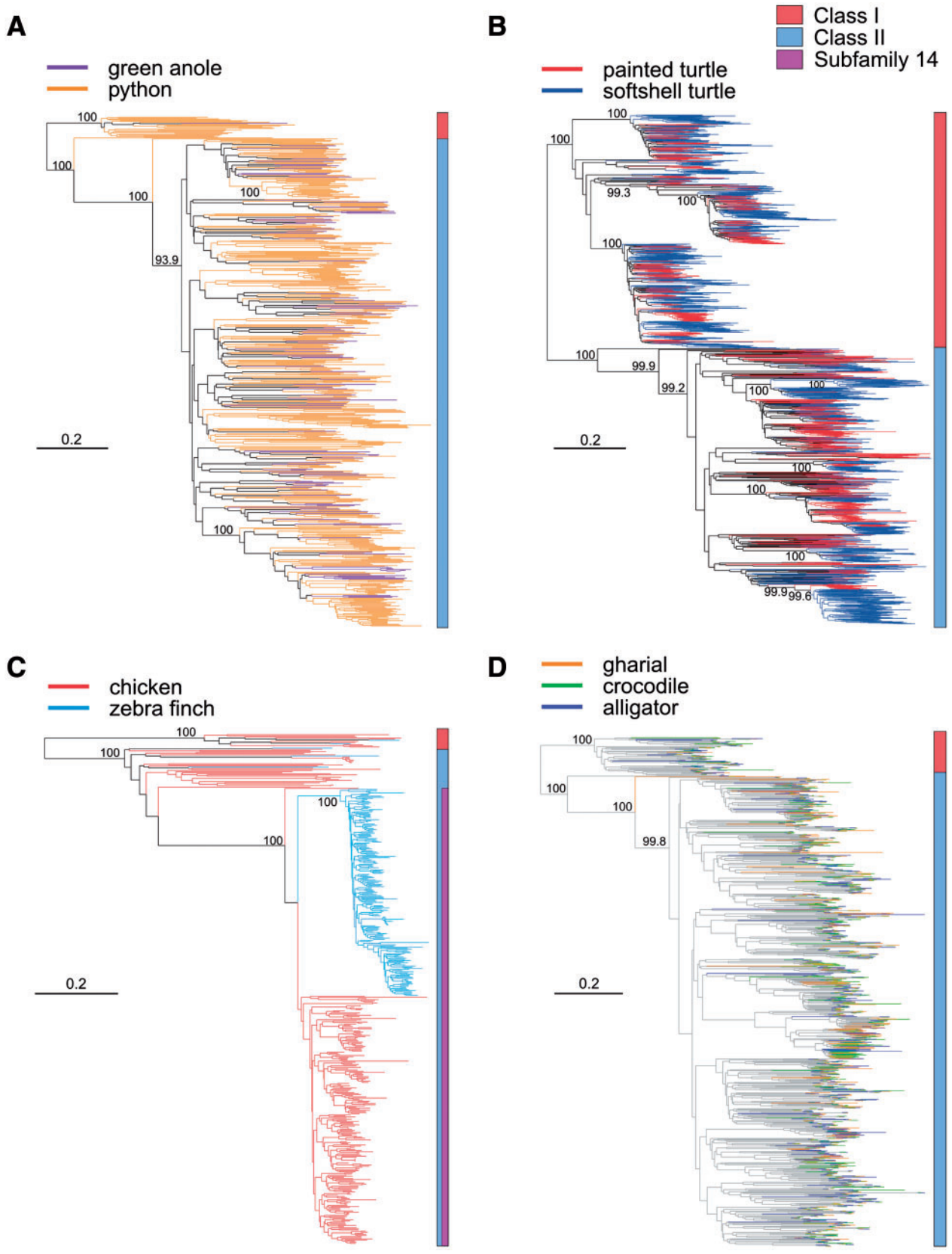

Fig. 5.—Independent phylogenetic reconstruction of ORs in each major group of saurposids. (*A*) squamates, (*B*) turtles, (*C*) birds, and (*D*) crocodilians.

number and diversity (fig. 5*D*). The conservative nature of crocodilian OR repertoires is exemplified by noticeably short terminal branches among orthologous gene copies (fig. 5*D*). These observations collectively suggest that not only have crocodilians not experienced substantial change in the number and diversity of OR genes, but have also experienced fewer amino acid changes among orthologous OR genes since extant crocodilians diverged from their common ancestor.

ORs Subfamilies in the Last Common Ancestor of Sauropsids

Consistent with previous analyses of OR repertoires in birds (Khan et al. 2015), our phylogenies indicate that at least six OR subfamilies (51, 52, 14, 4, 12, and 1/3/7) formed monophyletic groups among sauropsids, suggesting that these subfamilies began diversifying in the common ancestor of sauropsids. In addition, the majority of the predicted ORs in subfamilies 11, 10, 6, and 2/13 were placed in monophyletic groups as well, indicating that these subfamilies were also present in the last common ancestor of sauropsids. The most interesting case is subfamily 5/8/9, which is split into three weakly supported clades in our study (fig. 4*A*). Thus, our analyses suggest that the OR classification derived from mammals is largely applicable to sauropsids, and that the 11 major groups emerged prior to divergence of these mammals and sauropsids.

The Role of Natural Selection in Shaping OR Repertoires

The relative role selection played in shaping OR repertoires is a matter of debate. Early studies suggest that variation in OR repertoires is largely independent of selection (Niimura and Nei 2007; Nei et al. 2008). However, more recent comparative studies among mammals (Hayden et al. 2010, 2014) and birds (Khan et al. 2105) suggest that OR repertoires reflect ecological adaptations and have in part been shaped by natural selection. Khan et al. (2015) also show that olfactory acuity, as reflected by the size of the olfactory bulb, is correlated with the size of the olfactory repertoire.

Our results provide new and intriguing evidence consistent with a role of natural selection in shaping OR repertoires. We found independent expansions of subfamilies associated with detection of waterborne odorants in the two aquatic groups studied: subfamily 2/13 expanded in crocodiles, which has been linked to chemoreception in aquatic mammals and birds (Hayden et al. 2010; Khan et al. 2015), and Class I ORs in turtles, which are hypothesized to primarily bind waterborne odorants (Saito et al. 2009; Wang et al. 2013). Additional support for natural selection was observed in comparisons between squamate reptile OR repertoires. The green anole is an arboreal insectivore that relies on visual cues for social interactions (Leal and Fleishman 2004) and has the lowest number of functional ORs, despite having high OR subfamily diversity. In contrast, the python, which like most snakes, has poor hearing and vision and relies heavily on chemoreception to locate prey and mates, has at least five times as many putatively functional OR genes as the anole. Thus, the difference in size of squamate OR repertoires points toward a correspondence between OR size and the relative dependence on chemosensory information. While not conclusive, examples from sauropsid OR repertoires are at least consistent with natural selection playing a role in shaping OR repertoires, and suggests that the diversity of OR repertoires and natural history of sauropsid species may provide a rich model system for more detailed tests of this hypothesis.

Conclusions

Sauropsids represent an ecologically and phenotypically diverse set of tetrapods that include the closest living relatives to mammals, and recently available genomes of representative members of sauropsid lineages provide new opportunities to study the patterns of OR diversification in the group. Our results indicate that most sauropsids have diverse and relatively large OR repertoires that derive from a complex diversity of lineage-specific patterns of gene birth and death, and the differential retention of OR duplicates. We find that gene loss has played a prominent role in the evolution of the repertoires of birds and lizards. In contrast, turtles have experienced notable gains of class I ORs, and the common ancestor of crocodilians gained multiple ORs. Unlike other lineages, however, the crocodilian repertoire has remained nearly constant since the diversification of crocodilian lineages. Overall sauropsids have undergone numerous major life history and ecological transitions that are likely to have resulted in changes in the dependence of various lineages on olfaction and on OR repertoires.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the National Science Foundation [DEB-1355176 (F.G.H. and D.A.R.), EPS-0903787 (F.G.H.), MCB-0841821 (D.A.R.), and MCB-1052500 (D.A.R.)] as well as funding from the College of Agriculture and Life Sciences and the Institute for Genomics, Biocomputing, and Biotechnology at Mississippi State University.

Literature Cited

Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 65:175–187.

Castoe TA, et al. 2009. Dynamic nucleotide mutation gradients and control region usage in squamate reptile mitochondrial genomes. Cytogenet Genome Res. 127:112–127.

Castoe TA, et al. 2013. The Burmese Python genome reveals the molecular basis for extreme adaptation in snakes. Proc Natl Acad Sci U S A. 110:20645–20650.
Clark RW. 2007. Public information for solitary foragers: timber rattlesnakes use conspecific chemical cues to select ambush sites. Behav Ecol. 18:487–490.
Cooper WE Jr. 1991. Discrimination of integumentary prey chemicals and strike-induced chemosensory searching in the ball python, *Python regius*. J Ethol. 9:9–23.
De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22:1269–1271.
Dehara Y, et al. 2012. Characterization of squamate olfactory receptor genes and their transcripts by the high-throughput sequencing approach. Genome Biol Evol. 4:602–616.
Evans SE. 2003. At the feet of the dinosaurs: the early history and radiation of lizards. Biol Rev. 78:513–551.
Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB. 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol. 63:1256–1272.
Garrett EC, Steiper ME. 2014. Strong links between genomic and anatomical diversity in both mammalian olfactory chemosensory systems. Proc R Soc B Biol Sci. 281:20132828.
Giglio S, et al. 2001. Olfactory receptor–gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet. 68:874–883.
Glusman G, et al. 2000. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mamm Genome. 11:1016–1023.
Glusman G, Yanai I, Rubin I, Lancet D. 2001. The complete human olfactory subgenome. Genome Res. 11:685–702.
Green RE, et al. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346:1254449.
Hayden S, et al. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. Genome Res. 20:1–9.
Hayden S, et al. 2014. A cluster of olfactory receptor genes linked to frugivory in bats. Mol Biol Evol. 31:917–927.
Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971–2972.
Huerta-Cepas J, Dopazo J, Gabaldón T. 2010. ETE: a Python environment for tree exploration. BMC Bioinformatics 11:24.
Huerta-Cepas J, Gabaldón T. 2011. Assigning duplication events to relative temporal scales in genome-wide studies. Bioinformatics 27:28–45.
Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9:286–298.
Khan I, et al. 2015. Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. Mol Biol Evol. 32:2832–2843.
Kishida T, Hikida T. 2010. Degeneration patterns of the olfactory receptor genes in sea snakes. J Evol Biol. 23:302–310.
Kishida T, Kubota S, Shirayama Y, Fukami H. 2007. The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. Biol Lett. 3:428–430.
Lagerström MC, et al. 2006. The G protein-coupled receptor subset of the chicken genome. PLoS Comput Biol. 2:e54.
Leal M, Fleishman LJ. 2004. Differences in visual signal design and detectability between allopatric populations of Anolis lizards. Am Nat. 163:26–39.
LeMaster MP, Mason RT. 2002. Variation in a female sexual attractiveness pheromone controls male mate choice in garter snakes. J Chem Ecol. 28:1269–1285.
Matsui A, Go Y, Niimura Y. 2010. Degeneration of olfactory receptor gene repertories in primates: no direct link to full trichromatic vision. Mol Biol Evol. 27:1192–1200.
Mombaerts P. 1999. Seven-transmembrane proteins as odorant and chemosensory receptors. Science 286:707–711.
Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 9:951–963.
Niimura Y. 2009. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. Genome Biol Evol. 1:34–44.
Niimura Y, Matsui A, Touhara K. 2014. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome Res. 24:1485–1496.
Niimura Y, Nei M. 2003. Evolution of olfactory receptor genes in the human genome. Proc Natl Acad Sci U S A. 100:12235–12240.
Niimura Y, Nei M. 2005a. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. Gene 346:13–21.
Niimura Y, Nei M. 2005b. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc Natl Acad Sci U S A. 102:6039–6044.
Niimura Y, Nei M. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS One 2:e708.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.
Pyron RA. 2010. A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. Syst Biol. 59:185–194.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Roos J, Aggarwal RK, Janke A. 2007. Extended mitogenomic phylogenetic analyses yield new insight into crocodylian evolution and their survival of the Cretaceous–Tertiary boundary. Mol Phylogenet Evol. 45:663–673.
Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD. 2009. Odor coding by a mammalian receptor repertoire. Sci Signal. 2:ra9.
Shaffer HB, et al. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome Biol. 14:R28.
Shine R, Mason RT. 2001. Courting male garter snakes (*Thamnophis sirtalis parietalis*) use multiple cues to identify potential mates. Behav Ecol Sociobiol. 49:465–473.
Steiger SS, Fidler AE, Valcu M, Kempenaers B. 2008. Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? Proc R Soc B Biol Sci. 275:2309–2317.
Steiger SS, Kuryshev VY, Stensmyr MC, Kempenaers B, Mueller JC. 2009. A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group γ genes in birds. BMC Genomics 10:446.
Stone A, Holtzman DA. 1996. Feeding responses in young boa constrictors are mediated by the vomeronasal system. Anim Behav. 52:949–955.
Wan Q-H, et al. 2013. Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. Cell Res. 23:1091–1105.
Wang Z, et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat Genet. 45:701–706.
Zhang G, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.
Zhang X, Firestein S. 2002. The olfactory receptor gene superfamily of the mouse. Nat Neurosci. 5:124–133.

Associate editor: Naruya Saitou