Syst. Biol. 57(4):574–590, 2008
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150802304787

The Vestigial Olfactory Receptor Subgenome of Odontocete Whales: Phylogenetic Congruence between Gene-Tree Reconciliation and Supermatrix Methods

Michael R. McGowen,[1] Clay Clark,[1,2] and John Gatesy[1]

[1]*Department of Biology, University of California, Riverside, Riverside, California 92521, USA; E-mail: mmcgo002@student.ucr.edu (M.R.M.)*
[2]*Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA*

Abstract.—The macroevolutionary transition of whales (cetaceans) from a terrestrial quadruped to an obligate aquatic form involved major changes in sensory abilities. Compared to terrestrial mammals, the olfactory system of baleen whales is dramatically reduced, and in toothed whales is completely absent. We sampled the olfactory receptor (OR) subgenomes of eight cetacean species from four families. A multigene tree of 115 newly characterized OR sequences from these eight species and published data for *Bos taurus* revealed a diverse array of class II OR paralogues in Cetacea. Evolution of the OR gene superfamily in toothed whales (Odontoceti) featured a multitude of independent pseudogenization events, supporting anatomical evidence that odontocetes have lost their olfactory sense. We explored the phylogenetic utility of OR pseudogenes in Cetacea, concentrating on delphinids (oceanic dolphins), the product of a rapid evolutionary radiation that has been difficult to resolve in previous studies of mitochondrial DNA sequences. Phylogenetic analyses of OR pseudogenes using both gene-tree reconciliation and supermatrix methods yielded fully resolved, consistently supported relationships among members of four delphinid subfamilies. Alternative minimizations of gene duplications, gene duplications plus gene losses, deep coalescence events, and nucleotide substitutions plus indels returned highly congruent phylogenetic hypotheses. Novel DNA sequence data for six single-copy nuclear loci and three mitochondrial genes (>5000 aligned nucleotides) provided an independent test of the OR trees. Nucleotide substitutions and indels in OR pseudogenes showed a very low degree of homoplasy in comparison to mitochondrial DNA and, on average, provided more variation than single-copy nuclear DNA. Our results suggest that phylogenetic analysis of the large OR superfamily will be effective for resolving relationships within Cetacea whether supermatrix or gene-tree reconciliation procedures are used. [Cetaceans; Delphinidae; gene-tree reconciliation; mysticetes; odontocetes; olfactory receptors; pseudogenes; phylogeny; supermatrix.]

In most mammals, olfaction represents an essential chemosensory function necessary for survival. The molecular basis of olfaction resides in multiple G-protein–coupled receptors expressed in the sensory neurons of the olfactory epithelia. These olfactory receptors are short proteins characterized by seven transmembrane regions (Mombaerts, 2004) and are encoded by a large collection of single-exon genes that are distributed in clusters on multiple chromosomes (Glusman et al., 2001; Zhang and Firestein, 2002). Each functional olfactory receptor (OR) gene codes for a separate receptor that detects a specific set of odor molecules (Mombaerts, 2004).

OR genes compose the largest gene superfamily in mammalian genomes (Mombaerts, 2004; Niimura and Nei, 2005). There are ∼900 OR genes in *Homo sapiens* (Glusman et al., 2001; Niimura and Nei, 2003) as well as ∼1300 in both *Mus musculus* and *Canis familiaris* (Zhang and Firestein, 2002; Olender et al., 2004). In mammals, a large proportion of OR genes are nonfunctional: ∼20% of OR genes are pseudogenes in *Mus musculus* (Zhang and Firestein, 2002), ∼27% in *Canis familiaris* (Olender et al., 2004), and ∼52% to 63% in *Homo sapiens* (Glusman et al., 2001; Niimura and Nei, 2003). Rouquier et al. (2000) hypothesized that the proportion of functional OR genes in a species is roughly correlated with overall olfactory ability and anatomical complexity of the olfactory apparatus (olfactory bulb, olfactory nerve, cribriform plate). For example, in primates, a decrease in the proportion of functional OR genes is associated with reduction of olfactory anatomy in the lineage leading to humans (Rouquier et al., 2000).

Whales (cetaceans) represent a clade of mammals that have adapted to a purely aquatic lifestyle and as a result have a markedly reduced olfactory apparatus compared to their terrestrial cousins. Crown-group Cetacea is composed of two clades, toothed whales (Odontoceti) and baleen whales (Mysticeti), that differ considerably in the degree of this reduction. Olfactory structures are entirely lacking in adult odontocetes, implying a complete loss of the olfactory sense (Oelschläger, 1992). Mysticetes possess a highly reduced although intact olfactory apparatus, and anecdotal behavioral observations suggest that baleen whales might retain some sense of airborne smell (Cave, 1988; Oelschläger, 1992). At the molecular level, a small sample of OR gene sequences from three odontocete species (*Stenella coeruleoalba*, *Kogia sima*, and *Phocoenoides dalli*) revealed that an estimated 71% to 78% of OR genes were pseudogenes (Frietag et al., 1998; Kishida et al., 2007). A recent molecular survey suggested that the OR repertoire of one mysticete, *Balaenoptera acutorostrata* (minke whale), contains a smaller percentage of pseudogenes (58%) than odontocetes (Kishida et al., 2007).

Phylogenetic Utility of the OR Gene Superfamily

The OR subgenome of cetaceans offers the opportunity to examine the application of a large gene family to the reconstruction of species phylogeny and to assess the phylogenetic utility of pseudogenes. Multigene families can provide a wealth of character data that is often ignored a priori for fear of unrecognized paralogy (Martin and Burg, 2002). Undetected gene duplications can be a confounding source of conflict between a gene tree and its species tree due to the birth-and-death process of gene family evolution (Page and Charleston, 1997; Maddison, 1997). In spite of this perceived difficulty, there have been some successful applications of multigene families to phylogenetic reconstruction using a variety of methods (Goodman et al., 1979; Mathews and Donoghue, 2000; Carlini et al., 2000; Page, 2000; Cotton and Page, 2002; Martin and Burg, 2002).

Gene-tree reconciliation (GR) represents one proposed methodology that utilizes multigene families to reconstruct species phylogeny (Page and Charleston, 1997; Slowinski et al., 1997; Slowinski and Page, 1999). GR reconciles one or more gene trees with an underlying species phylogeny by minimizing the cost from a weighted sum of inferred gene duplications and losses, gene conversions, introgression events, lateral transfers, and deep coalescences. However, simultaneous minimizations of duplications/losses and deep coalescence events may not be feasible without additional biological information (J. Cotton, personal communication), and optimizations of gene conversions, lateral transfers, and introgression events have yet to be implemented in a general, comprehensive framework.

Alternatively, paralogy relationships can be hypothesized by direct sequence comparison and phylogenetic analysis (Mathews and Donoghue, 2000; Carlini et al., 2000). Using this approach, putative orthologue groups can be isolated and then concatenated into a single supermatrix in which each taxon is represented in the data set only once (Carlini et al., 2000; Simmons et al., 2000). However, several authors have argued that the supermatrix method does not directly take into account the distinct historical patterns for different alleles, genes, and gene duplicates (Bull et al., 1993; Miyamoto and Fitch, 1995; Huelsenbeck et al., 1996; Slowinski and Page, 1999; Kubatko and Degnan, 2007). The OR subgenome represents an extreme among mammalian multigene families in terms of the overall number of paralogues present. Therefore, comparative analysis of OR genes is ideally suited for evaluating the relative merits and shortcomings of these systematic methodologies. Parallel supermatrix and GR analyses of the same database have been rare (Mathews and Donoghue, 2000; Simmons and Freudenstein, 2002; Cotton and Page, 2003).

The presence of several hundred OR pseudogenes in a mammalian lineage is also a potential boon for molecular systematists who seek rapidly evolving nuclear (nu) DNA sequences for analysis. OR genes are distributed over many linkage groups (Zhang and Firestein, 2002) and are easily amplified using degenerate PCR primers (Freitag et al., 1998; Rouquier et al., 2000; Gilad et al., 2004). Furthermore, OR genes show increased rates of both base substitution and indel accumulation after pseudogene formation (Whinnett and Mundy, 2003); the rate of nucleotide substitution in pseudogenes is thought to be approximately equal to the overall nuclear rate of spontaneous mutation (Gojobori et al., 1982). Thus, pseudogenes generally would be expected to diverge at a higher rate than nuclear protein-coding regions, introns, and regulatory sequences, which are constrained by negative selection (Li et al., 1981; Gojobori et al., 1982; Ophir and Graur, 1997).

Highly variable OR pseudogenes may assist in resolving problematic relationships within Odontoceti, especially among major lineages of the family Delphinidae (oceanic dolphins). Fossil evidence suggests that oceanic dolphins experienced a rapid radiation in the Late Miocene or Pliocene (∼3 to 12 Ma; Barnes, 2002), resulting in ∼35 closely related extant species in 17 genera (LeDuc et al., 1999). Mitochondrial (mt) DNA sequences have been the primary focus for molecular systematic studies of delphinids and other recently diverged mammalian species (Brown et al., 1982; Irwin et al., 1991; Allard et al., 1992). Phylogenetic analyses of mtDNA were successful in grouping most dolphin species into major clades but generally did not resolve short branches at the base of the tree with robust support (Milinkovitch et al., 1994; LeDuc et al., 1999; Hamilton et al., 2001; Harlin-Cognato and Honeycutt, 2006; May-Collado and Agnarsson, 2006). Because of its high mutation rate, mtDNA may quickly become saturated by multiple overlapping substitutions, hindering resolution of rapid evolutionary radiations where short internodes are numerous. Additionally, the close linkage of all mitochondrial genes limits the phylogenetic independence of different mitochondrial loci (Kraus and Miyamoto, 1991; Allard et al., 1992; Machado and Hey, 2003).

Previous studies have shown that extensive nuDNA sequence data can offset these confounding issues (Matthee and Davis, 2001; Matthee et al., 2001; Steppan et al., 2005). In mammals, nuDNA generally evolves at a slower rate than mtDNA, yielding systematic characters with much lower levels of homoplasy relative to mtDNA (Gatesy et al., 1996; Matthee and Davis, 2001; Matthee et al., 2001; Springer et al., 2001), but very large nuDNA data sets are necessary to discern recent divergence events (Steppan et al., 2005). We predicted that nuclear OR pseudogenes would robustly resolve basal relationships of oceanic dolphins better than saturated mitochondrial genes. Furthermore, because OR pseudogenes have been freed from selection, these sequences should provide more informative character variation than typical nuDNA markers (i.e., introns and exons) that are constrained by negative selection.

The theoretical advantages (and disadvantages) of supermatrix methods and supertree methods (including GR) have been outlined in recent reviews (Bininda-Emonds, 2004a; de Queiroz and Gatesy, 2007). Here, we apply these alternative modes of analysis to a large comparative database for odontocete whales and outgroups (11 putative OR orthologue groups and 10 additional mitochondrial and nuclear genes). We use multiple GR and supermatrix analyses of the OR subgenome to contrast these different systematic approaches and record phylogenetic congruence and conflict across methods. We estimate the relative rate of sequence divergence in OR pseudogenes and compare the phylogenetic utility of these markers in comparison to independent mtDNA and nuDNA data, especially with reference to the rapid radiation of delphinid dolphins. In addition, we characterize the diversity of class II OR genes in cetaceans, further document degradation of the OR subgenome in

Odontoceti, and relate this mass pseudogenization to the parallel loss of the olfactory system in multiple odontocete lineages.

METHODS

Taxa and Samples

Taxa were chosen to represent Mysticeti and three divergent clades within Odontoceti (Physeteridae, Phocoenidae, and Delphinidae). We obtained DNA samples of seven cetacean species from the Southwest Fisheries Science Center (SWFSC; La Jolla, California) and one species (*Physeter macrocephalus*) from Dr. M. Milinkovitch (Yale University; presently Free University of Brussels). Species and samples used in this study were Mysticeti, *Eubalaena japonica* (North Pacific right whale, SWFSC no. Z13190); Physeteridae, *Physeter macrocephalus* (sperm whale); Phocoenidae, *Phocoena phocoena* (harbor porpoise, SWFSC no. Z28452 [from Marine Mammal Center, Sausalito, CA], no. Z32); Delphinidae, *Orcinus orca* (killer whale, SWFSC no. Z6004 [from Marine Mammal Center, Sausalito, CA]), *Pseudorca crassidens* (false killer whale, SWFSC no. Z38069), *Steno bredanensis* (rough-toothed dolphin, SWFSC no. Z38282), *Delphinus delphis* (short-beaked common dolphin, SWFSC no. Z31912), and *Stenella coeruleoalba* (striped dolphin, SWFSC no. Z37941). The five species sampled from Delphinidae exemplify four deep lineages within the group according to analyses of cytochrome *b* (LeDuc et al., 1999; May-Collado and Agnarsson, 2006).

PCR Amplification, Cloning, and Sequencing of OR Genes

For each species, multiple OR genes were amplified simultaneously in a 50-μL reaction mixture containing cetacean genomic DNA, 67 mM Tris, 3 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 200 μM dNTPs, 2 μM of each primer, and 0.75 U of *Taq* polymerase (Invitrogen, Carlsbad, CA). Polymerase chain reaction (PCR) amplifications included an initial denaturation phase at 94°C (2 min), then 50 cycles at 94°C (1 min), 50°C (1 min), 72°C (1 min), and a final elongation phase at 72°C (2 min). Primers for the initial amplification of OR gene fragments were OR 2.3 and OR 6.1 from Freitag et al. (1998). These are degenerate sequences from transmembrane domains 2 and 6, respectively, of class II ORs. PCR products were purified using Montage PCR Centrifugal Filter Devices (Millipore, Bedford, MA) and then cloned using the pCR 4-TOPO vector (Invitrogen). A PCR procedure similar to the one described above, but with annealing temperature set to 55°C to 58°C, was used to amplify individual colonies using the universal primers M13F and M13R. At least 32 clones were sequenced for each species using an ABI PRISM 3730xl DNA Sequencer (Applied Biosystems). Sequences were deposited in GenBank (EU684973 to EU685087).

Orthologue Identification

Freitag et al. (1998) identified 17 OR sequences from *Stenella coeruleoalba*, one of our sampled taxa. Twelve of these OR gene fragments (GenBank AJ233789 to AJ233793, AJ233799 to AJ233805) span the region that was PCR amplified here and were included with our new data (115 sequences) in comparative analyses. BLASTN searches (Altschul et al., 1990) against the NCBI database were executed for each of the above 127 OR sequences. To keep the number of sequences to a manageable level, only the top five matches for each cetacean sequence were held. If the top five matches did not include an OR sequence from *Homo sapiens* or *Canis familiaris*, the top match for each of these species also was downloaded, provided they were not lower than the 10th best match. All matches retained were significant with E values $<1 \times 10^{-4}$. Cetacean sequences also were compared to the *Bos taurus* (domestic cow) genome (build 2.1) using BLASTN. *Bos taurus* is included, along with cetaceans, in the mammalian order Cetartiodactyla (Montgelard et al., 1997) and represents the closest relative to Cetacea with a genomic assembly. All matches to the cow genome with alignment scores $\geq$200 bits were downloaded and saved for further analysis; all of these possessed E values $<1 \times 10^{-78}$.

To place our OR sequences relative to those from other mammalian species and to clarify orthology/paralogy relationships, cetacean ORs were aligned with the top matches from BLASTN searches (see above); these 319 OR sequences were aligned using Clustal W (Thompson et al., 1994) as implemented in MacVector 7.2.3 (Accelrys). Gap-opening penalty was set at 10, gap extension penalty was 1, and default settings were used for other alignment parameters. A heuristic parsimony analysis of the 319 sequences was conducted in PAUP* (Swofford, 2002; 100 random taxon addition replicates with TBR branch swapping). Paralogy was hypothesized based on clustering of the cetacean OR sequences with annotated mammalian OR genes (following Simmons et al., 2000). BLASTN searches against the HORDE mammalian OR database (bioportal.weizmann.ac.il/HORDE/; Glusman et al., 2000, 2001) were used to confirm OR family and subfamily identities implied by phylogenetic analysis, and names were applied to cetacean sequences based on closest matches to *Homo sapiens* and *Canis familiaris* sequences. The classification system for OR genes outlined in Glusman et al. (2000) was utilized. In this framework, OR genes are named by what family they belong to (a number), followed by subfamily identification (one or two letters), and ending in another number differentiating between members of the same subfamily.

Pseudogenes and Optimization of Nonsense Mutations

Pseudogenes were identified by the presence of frameshift mutations in DNA sequence alignments and by premature stop codons in translated sequences (see Freitag et al., 1998). To effectively screen the OR subgenome of cetaceans for phylogenetic information, degenerate PCR primers that amplify ~55% of the coding sequence were utilized. Because the entire protein-coding region was not amplified, some sequences that were classified as functional OR genes might, in actuality, be non-functional pseudogenes. Therefore, our estimated percentages of non-functional versus functional OR genes are likely underestimates given this conservative procedure. OR subfamily alignments (see below) were used to map frameshift indels/nonsense mutations on gene trees for different orthologue groups. Indel characters and base substitutions were optimized onto gene trees by parsimony using PAUP*; the polarities of mutational changes within Cetacea were determined by outgroup comparisons to closely related *Bos* sequences.

To estimate the relative divergence of OR pseudogenes within Delphinidae, we aligned sequences from *Orcinus orca* (Orcininae) and *Delphinus delphis* (Delphininae), two morphologically disparate species. Comparisons were made between OR pseudogenes (nine putative loci) and other partitions of DNA sequence data: mitochondrial D-Loop (537 base pairs [bp]), mitochondrial protein-coding genes (*MT-CYB*: 1140 bp; *MT-CO2*: 684 bp), mitochondrial ribosomal (r) DNAs (12S rDNA: 404 bp; 16S rDNA: 546 bp), single-copy nuclear introns (*PRM1*: 98 bp; *SPTBN1*: 657 bp; *BTN1A1*: 366 bp; *LALBA*: 552 bp; *ACTB*: 813 bp), single-copy nuclear exons (*PRM1*: 144 bp; *RAG1*: 798 bp; *AMBN* : 467 bp; *RAG2*: 474 bp; *SPTBN1*: 217 bp; *SRY*: 606 bp; *TBX4*: 1336 bp). For each partition listed above, a pairwise distance between the sequences for these two species was calculated using the general time-reversible (GTR) model of evolution as implemented in PAUP*. A separate analysis using uncorrected pairwise distances obtained similar results (not shown). Number of indels per 1000 bp was also determined for each gene class listed.

Sequence Alignment

To examine evolutionary relationships among cetacean OR genes in a more manageable context, we generated alignments using subsets of our initial compilation of 319 mammalian OR genes. First, all 127 cetacean OR sequences (115 new + 12 published) were aligned in Clustal W using parameters described above. Gaps in the resulting alignment were coded as separate binary characters using the simple gap-coding procedure of Simmons and Ochoterena (2000) that is automated in the "Indel Coder" feature of SeqState v.1.25 (Müller, 2005). The final data matrix (127 cetacean sequences plus indels) is referred to as the "Cetacea-only" matrix in the remainder of the paper.

Next, putative OR orthologue groups that included sequences from three or more cetacean species were identified. The sequences from these groups were aligned with *Bos taurus* sequences using Clustal W (see parameters above); for each orthologue group, *Bos* sequences were the closest available outgroup sequences. Gaps in the alignment were coded using SeqState as above. The final data set of 80 cetacean sequences, 18 bovine sequences, and associated indel characters was retained for subsequent phylogenetic analyses and in the remainder of this paper is referred to as the "Cetacea + *Bos*" matrix.

Finally, each hypothesized OR orthologue group represented by three or more cetacean species was aligned separately without interference from highly divergent OR paralogues. *Bos taurus* sequences were included in these alignments so that the polarity of character state changes could be estimated. In total, 11 separate Clustal W alignments were executed, and gap characters were coded as above. These alignments were retained for phylogenetic analysis and subsequent optimization of indel characters.

Independent Mitochondrial and Nuclear DNA Evidence

To evaluate OR trees using independent data, we compiled sequences from four mitochondrial genes and six single-copy nuclear genes; these genes are a subset of those used in the pairwise comparisons described above. Mitochondrial genes were cytochrome *b* (*MT-CYB*: 1140 bp), cytochrome *c* oxidase II (*MT-CO2*: 684 bp), and partial sequences from 12S rDNA (407 bp + 9 gap characters) and 16S rDNA (553 bp + 10 gap characters). Nuclear markers included *RAG1* (exon; 798 bp), *PRM1* (exons + intron + flanking sequences; 430 bp + 10 gap characters), *LALBA* (exon + intron; 565 bp + 11 gap characters), *SPTBN1* (exons + intron; 878 bp + 9 gap characters), *BTN1A1* (intron; 367 bp + 6 gap characters), and *AMBN* (exon; 471 bp + 2 gap characters). This compilation consisted of a total of 90 sequences. Thirty-two were downloaded from Genbank; the remaining 58 were PCR amplified and sequenced using methods in Deméré et al. (2008). PCR primers are listed in Table 1. GenBank accession numbers for published and newly generated sequences are listed in Table 2. Sequence alignment was conducted as above.

TABLE 1. Primer sequences for mitochondrial (mt) DNA and single-copy nuclear genes. All primers are 5′ to 3′. For *BTN1A1*, BTR219F was used for PCR amplification, and BTR232F was used for sequencing.

12S rDNA	Milinkovitch et al., 1994
16S rDNA	Milinkovitch et al., 1994
MT-CO2	This study
	Forward CO2LCet: TAAARTCTTACATAACTTTGTC
	Reverse CO2RCet: TCTCAATCTTTAACTTAAAAGG
RAG1	Deméré et al., 2008
PRM1	Queralt et al., 1995; Deméré et al., 2008
LALBA	Milinkovitch et al., 1998
BTN1A1	This study
	Forward BTR219F: GGAGATGAGTAGGAAGGGGGTTTG
	BTR232F: GGTTTGAGTTGASAGTG
	Reverse BTR585R: TGGCTTGAAAGGAAAAAGGAAAC
AMBN	Deméré et al., 2008
SPTBN1	Fragment 1:
	SPTBN1CF, GAAGACCTGTTACAGAAGCA (Matthee et al., 2001)
	SPTBN1R460A, TTTTGATCACTTAGGAACCA (This study)
	SPTBN1R460B, TTTTGATCACTTGGGAACGA (This study)
	Fragment 2:
	SPTBN1F570, TCCCTCCTCATCCAGTCAAG (This study)
	SPTBNBR, CTGCCATCTCCCAGAAGAA (Matthee et al., 2001)

TABLE 2. GenBank accession numbers for mtDNA genes and single-copy nuclear (nuDNA) genes. In three cases, published sequences of close relatives were equated with species in our data set.

	mtDNA			
	MT-CYB	12S rDNA	16S rDNA	*MT-CO2*
Bos taurus	DQ124379	DQ124379	DQ124379	DQ124379
Eubalaena japonica	AP006474	AP006474	AP006474	AP006474
Physeter macrocephalus	AJ277029	AJ277029	AJ277029	AJ277029
Phocoena phocoena	AJ554063	AJ554063	AJ554063	AJ554063
Orcinus orca	AF084060	EU685088*	EU685093*	EU685098*
Pseudorca crassidens	AF084057	EU685089*	EU685094*	EU685099*
Steno bredanensis	AF084077	EU685090*	EU685095*	EU685100*
Delphinus delphis	AF084084	EU685091*	EU685096*	EU685101*
Stenella coeruleoalba	AF084082	EU685092*	EU685097*	EU685102*
	nuDNA			
	RAG1	*PRM1*	*LALBA*	
Bos taurus	NW928953	M18395	X06366	
Eubalaena japonica	EU445025	EU444939	(*E. australis* AY398660)	
Physeter macrocephalus	EU445013	EU444927	AF304098	
Phocoena phocoena	EU697424*	EU697404*	AJ007811	
Orcinus orca	EU697425*	EU697405*	EU697399*	
Pseudorca crassidens	EU697426*	EU697406*	EU697400*	
Steno bredanensis	EU697427*	EU697407*	EU697401*	
Delphinus delphis	EU697428*	EU697408*	EU697402*	
Stenella coeruleoalba	EU697429*	EU697409*	EU697403*	
	BTN1A1	*AMBN*	*SPTBN1*	
Bos taurus	AF037402	AF157019	AF165718	
Eubalaena japonica	EU697416*	EU445000	(*Balaena mysticetus* AF165638)	
Physeter macrocephalus	EU697417*	EU445004	(*Kogia breviceps* AF165646)	
Phocoena phocoena	EU697418*	EU697410*	EU697393*	
Orcinus orca	EU697419*	EU697411*	EU697394*	
Pseudorca crassidens	EU697420*	EU697412*	EU697395*	
Steno bredanensis	EU697421*	EU697413*	EU697396*	
Delphinus delphis	EU697422*	EU697414*	EU697397*	
Stenella coeruleoalba	EU697423*	EU697415*	EU697398*	

*This study.

Standard Phylogenetic Analyses

We executed maximum parsimony (MP), maximum likelihood (ML), and Bayesian analyses of the "Cetacea + *Bos*" and "Cetacea-only" OR matrices. Parsimony searches using PAUP* 4.0b10 (Swofford, 2002) were heuristic with at least 100 random stepwise-addition replicates and tree bisection reconnection (TBR) branch-swapping. All nucleotide substitutions and indel events were given equal weight, and internal branches were collapsed if the minimum length of an internode was zero ("amb-" option in PAUP*). Strict consensus trees were used to summarize relationships supported by all equally parsimonious topologies. Support for nodes was evaluated by non-parametric bootstrapping using 1000 pseudoreplicates (Felsenstein, 1985).

Maximum likelihood analyses were performed using PhyML (Guindon and Gascuel, 2003) with the GTR+I+Γ model as chosen by the Akaike information criterion (AIC) in ModelTest v3.6 (Posada and Crandall, 1998; Posada and Buckley, 2004). Starting trees were generated via the neighbor-joining method, and ML bootstrap support scores were computed using 100 pseudoreplicates.

Markov chain Monte Carlo (MCMC) Bayesian analyses were conducted using default parameters in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) and four simultaneous chains (three "cold"and one "heated"). MrModelTest 2.2 (Nylander, 2004) was used to choose optimal models according to AIC (Table 3), and the binary model was employed for gap characters (Ronquist et al., 2005). Two separate runs of 20,000,000 generations were employed for each matrix with trees sampled every 100 generations. Output parameters from Bayesian analyses were visualized using Tracer v.1.4 (Rambaut and Drummond, 2007) to ascertain stationarity and whether the tandem runs had converged on the same mean likelihood. In all cases, analyses reached stationarity before 15 million generations; output trees from the first 15 million generations were discarded as burn-in. Standard deviation of split frequencies in the last 5 million generations of both analyses was <0.013. A 50% majority-rule consensus of the remaining trees was taken to summarize posterior probabilities for each clade.

In the absence of an outgroup, all MP, ML, and Bayesian consensus trees for the "Cetacea + *Bos*" and "Cetacea only" matrices were midpoint rooted. Because of the shortcomings of midpoint rooting and its assumption of a molecular clock, we also tested the placement of the root by gene tree reconciliation using the program Notung 2.5 (Durand et al., 2006). Notung minimizes the weighted number of gene duplications and losses for each possible root (i.e., every internode) of a gene tree given a certain species tree. Here we input the species

TABLE 3. Models used in maximum likelihood (ML) and/or Bayesian analyses.

"Cetacea only"	GTR+I+Γ
"Cetacea + *Bos*"	GTR+I+Γ
OR1I1	HKY
OR2AT1	GTR
OR6M1	HKY+Γ
OR10A1	HKY
OR10AB1	GTR+Γ
OR10J1	GTR+Γ
OR10J2	GTR
OR10K1	GTR+Γ
OR10K3	HKY
OR13F1	HKY+I
OR13J1	HKY
MT-CYB	GTR+I+Γ
MT-CO2	GTR+I+Γ
12S rDNA	GTR+I+Γ
16S rDNA	GTR+I+Γ
AMBN	GTR
BTN1A1	GTR
LALBA	HKY+Γ
PRM1	HKY+Γ
RAG1	HKY+Γ
SPTBN1	HKY+Γ

tree derived from our GeneTree analyses (see below). In all cases, the midpoint root was one of many equally parsimonious roots, and all equally parsimonious roots were outside identified orthologue groups.

Separate parsimony and Bayesian analyses were also conducted as above for each putative linkage group: mtDNA (with separate genes partitioned in the Bayesian analysis), each of the 11 OR orthologue groups, and each single-copy nuclear gene. Gaps were coded for each data set as above, and topologies were rooted with sequences from *Bos taurus.* Matrices for all analyses were deposited in TreeBASE (Accession no. SN3957).

Supermatrix Analyses

The 11 separate OR orthologue group matrices were concatenated to form a single supermatrix data set. In this combined matrix, each of the included species (eight cetaceans + *Bos taurus*) was represented only once, and alignments for each orthologue group were placed end to end, resulting in a data set of nine taxa and 5755 characters (5719 nucleotides and 36 indels). In cases where two or more OR genes from the same species clustered together or were not resolved relative to other species, multiple sequences were represented using ambiguity codes to encompass all variation at specific sites within a species. Ambiguity codings were interpreted as representations of allelic heterogeneity, ambiguity from cloning error, or variation due to very recent gene duplication. Putative *OR13J1* alleles from *Stenella coeruleoalba* did not cluster; separate supermatrix analyses were conducted with each alternative sequence included. MP analyses were as described above. Bayesian analyses were performed with 12 data partitions (11 separate orthologue groups + gap characters). Models for sequence data in each analysis were based on the AIC output of MrModelTest 2.2 (Table 3). The binary model was utilized for indel characters. All data partitions were unlinked to permit independent estimation of model parameters among orthologue groups.

Data also were concatenated to form three additional combined data sets: (1) all mtDNA, (2) all single-copy nuDNA, and (3) a total combined supermatrix (mtDNA + OR genes + single-copy nuDNA). MP and Bayesian analyses were conducted as above; for Bayesian analyses, data sets were partitioned by gene (Table 3). The *OR13J1B* sequence for *Stenella coeruleoalba* was utilized in the total combined supermatrix.

Gene-Tree Reconciliation Analyses

Species trees were generated from gene trees using gene-tree reconciliation (GR) as implemented in the program GeneTree 1.3.0 (Page, 1998). Algorithms in GeneTree 1.3.0 evaluate the minimization of gene duplication events, gene duplications + losses, or deep coalescences on gene trees over all possible species trees. GeneTree computes the number of these events by fitting the gene tree to each species tree; the species tree with the lowest number of duplications/losses or deep coalescent events is considered optimal. As stated above, GeneTree cannot simultaneously account for gene duplication + loss and deep coalescence events; both processes can cause disagreements between gene and species trees (Page and Charleston, 1997; Maddison, 1997).

Here, we conducted two sets of GR analyses. The first group of analyses minimized costs due to duplications or duplications + losses in the OR superfamily tree derived from the "Cetacea + *Bos*" matrix. The GeneTree program only accommodates input trees that are fully bifurcating and rooted. To incorporate branch support in GeneTree analyses, a set of parsimony boostrap trees is used as an input (Cotton and Page, 2003). Unfortunately, rigorous parsimony bootstrap searches of the large "Cetacea + *Bos*" data set were not computationally feasible without the "amb-" option in effect (in PAUP*, this command collapses zero length branches and greatly speeds up tree search). Therefore, for this set of GR analyses, we utilized 100 trees drawn at random from the posterior probability distribution (post burn-in) from the Bayesian analysis of the "Cetacea + *Bos*" data set. Cotton and Page (2003) suggested that the use of a set of Bayesian trees might be more statistically rigorous than parsimony bootstrap trees. For each of these 100 OR superfamily topologies, optimal species trees were generated from a heuristic GR search with 10 random taxon addition replicates and alternating NNI and SPR branch swapping. The combined set of species trees, derived from these 100 heuristic searches, was summarized by a weighted majority-rule species tree ("gene-tree bootstrapping" option). This approach to gene-tree analysis accounts for the relative strength of support for different relationships in the overall multigene tree (Cotton and Page, 2002, 2003). For example, a node found in all 100 input multigene trees would influence the final majority-rule species tree more than a node supported by only 40 of the 100 input trees. Two separate GR analyses were performed: (1) minimization of gene duplications for the entire OR superfamily tree, and (2) minimization of gene duplications + gene losses for the entire OR superfamily tree.

The second set of GR analyses minimized costs due to deep coalescence events. These analyses were conducted using input gene trees derived from either parsimony bootstrap or Bayesian analyses of each individual OR orthologue group (in total, 11 putative OR loci), the combined mtDNA data, and each single-copy nuclear gene (in total, six nuclear loci). In the parsimony analyses, 100 trees were generated for each OR locus and for each of the independent mtDNA and nuDNA data sets. For each data set, a single bifurcating tree was saved from each of 100 parsimony bootstrap replicates. In the Bayesian analyses, 100 trees were drawn at random from the posterior probability distribution (post-burn-in) for each OR locus and for each of the independent mtDNA and nuDNA data sets. Four GR analyses were performed with a minimization of deep coalescences in effect using the "gene-tree bootstrapping" procedure outlined in Cotton and Page (2003): (1) the 11 OR genes with 100 input trees for each gene derived from parsimony bootstrap analyses; (2) the 11 OR genes with 100 input trees for each gene derived from Bayesian analyses; (3) the 11 OR genes + mtDNA + 6 single-copy nuclear genes with input trees derived from parsimony bootstrap analysis for each locus; and (4) the 11 OR genes + mtDNA + 6 single-copy nuclear genes with 100 input trees for each locus derived from Bayesian analyses. All GR searches were done in GeneTree 1.3.0 using the search parameters described above, so that the relative support for nodes among input trees was accounted for. Weighted majority-rule consensus trees were used to summarize the output of the GR analyses.

Results and Discussion

Diversity and Phylogeny of Cetacean OR Genes

Our PCR screen revealed a high diversity of OR genes in Cetacea; 115 distinct sequences from 309 clones were recovered. Phylogenetic analysis of the cetacean sequences and their top five BLASTN hits estimated the presence of at least 48 OR orthologue groups. The OR superfamily has been divided into families and subfamilies of genes based on sequence divergence (Glusman et al., 2000). BLAST searches indicated that the 115 cetacean sequences from this study, plus the 12 sequences included from Freitag et al. (1998), were distributed among a diverse array of class II OR families (families 1 to 2 and 4 to 13), with the majority (48) derived from family 10. Sequences ranged from 491 to 535 bp in length; most homologous fragments from *Bos taurus* were 518 bp (the primitive, presumably functional state). Based on comparisons of each cetacean orthologue group with homologous *Bos taurus* sequences and more distant outgroups, there were at least 73 indel events within Cetacea. The much higher frequency of inferred deletions (62) relative to insertions (11) was consistent with previous studies (de Jong and Ryden, 1981; Ophir and Graur, 1997). Deletions ranged in size from 1 to 24 bp, and insertions were 1 to 20 bp in length; 10 indels were multiples of 3 bp and did not result in frameshifts (nine deletions and one insertion).

Phylogenetic analyses of the 127 cetacean sequences ("Cetacea-only" matrix) showed a high level of divergence among sequences representing different OR gene families and subfamilies (Fig. 1). All OR subfamilies, however, were well supported by bootstrap scores and Bayesian posterior probabilities. Some members of OR families did not cluster (OR5, OR8, OR9, OR10, OR13), but for some families, this may be attributed to the use of pairwise genetic distances for demarcation of OR families in current classifications (see Niimura and Nei, 2003). Relationships among cetaceans differed across orthologue groups. Differences among MP, ML, and Bayesian results mainly concerned relationships of OR families and subfamilies (Fig. 1). Low support scores and differences in topology were unsurprising at this level, because subfamilies often were separated by branch lengths greater than 0.5 substitutions per site and may have diverged before the origin of mammals (Niimura and Nei, 2003).

Phylogenetic analysis of the "Cetacea + *Bos*" matrix included the most taxonomically well-represented OR subfamilies (Fig. 2). ML and Bayesian consensus trees differed from the MP tree at one deep node connecting different OR subfamilies. As in the "Cetacea-only" analyses, relationships among species were not wholly congruent across orthologue groups; for example, subgroupings within Delphinidae varied from gene to gene. In most cases, multiple sequences from a species that were assigned to the same orthologue group (putative alleles) were monophyletic or were not resolved relative to sequences from other species. There were, however, two exceptions. In *OR1I1*, *Stenella coerueloalba* sequences did not form a clade in the Bayesian tree, but they clustered in both MP and ML trees. The other exception was *OR13J1*; two sequences from *S. coeruleoalba* differed at 11 of 510 sites and did not form a clade to the exclusion of other OR sequences (Fig. 2). This pattern implies retention of ancestral polymorphism, recent gene duplication, or interspecies hybridization.

Pseudogene Content of OR Repertoires and the Degradation of Olfaction in Odontocetes

Translation of 115 cetacean OR genes showed that 84 encoded at least one stop codon that interrupted the reading frame. Comparisons with outgroup sequences demonstrated that in most cases the stop codon resulted from an upstream indel. Another six sequences had gaps near their 3′ ends, resulting in frame shifts with no observable stop codons in the sequenced region. However, such indels would cause shifts in the remaining downstream codons (~80 triplets), resulting in potentially nonfunctional gene products. Taking these together, 90 of 115 sequences (78.3%) were estimated to be pseudogenes, according to our conservative criteria

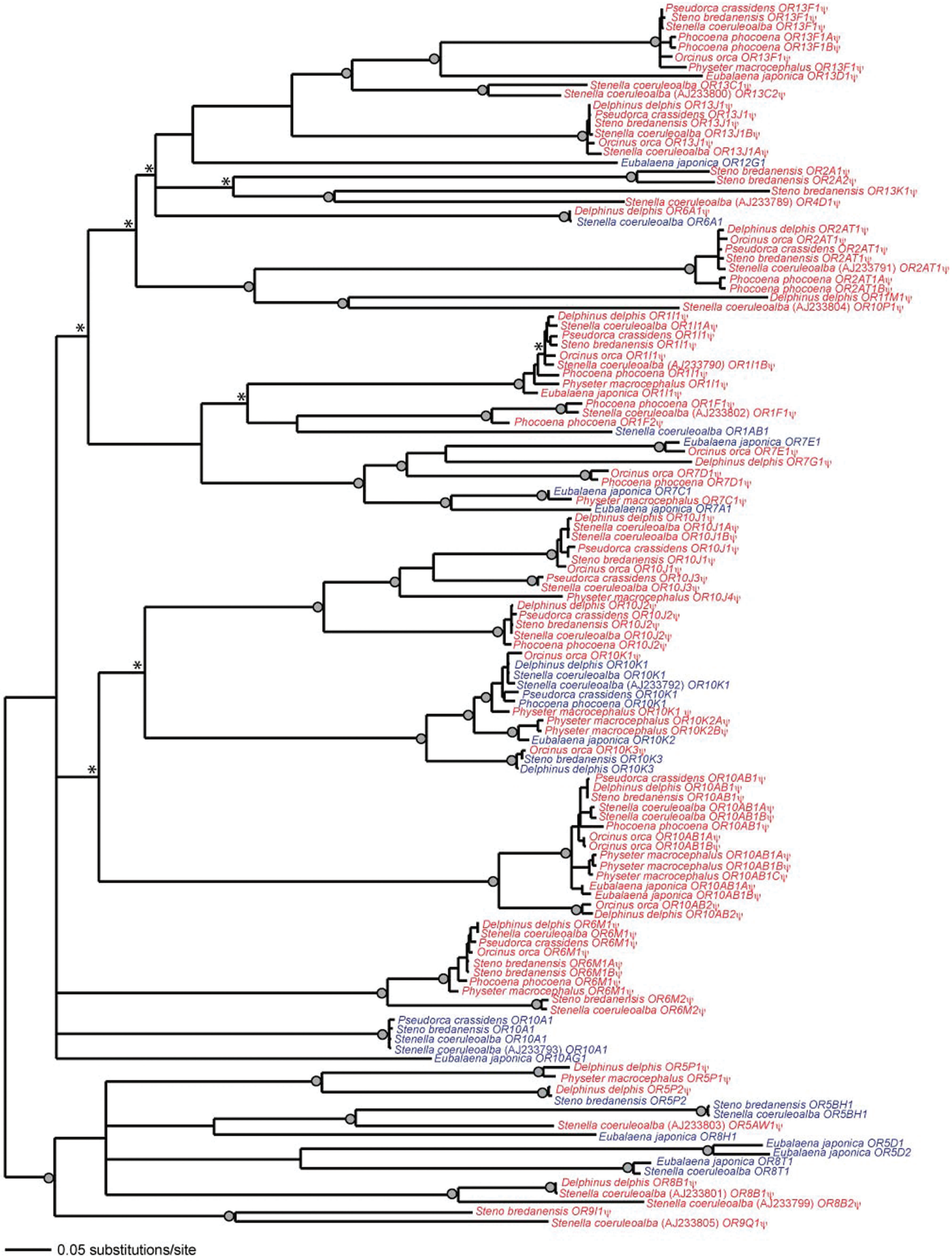
Pseudorca crassidens OR13F1ψ
Steno bredanensis OR13F1ψ
Stenella coeruleoalba OR13F1ψ
Phocoena phocoena OR13F1Aψ
Phocoena phocoena OR13F1Bψ
Orcinus orca OR13F1ψ
Physeter macrocephalus OR13F1ψ
Eubalaena japonica OR13D1ψ
Stenella coeruleoalba OR13C1ψ
Stenella coeruleoalba (AJ233800) OR13C2ψ
Delphinus delphis OR13J1ψ
Pseudorca crassidens OR13J1ψ
Steno bredanensis OR13J1ψ
Stenella coeruleoalba OR13J1Bψ
Orcinus orca OR13J1ψ
Stenella coeruleoalba OR13J1Aψ
Eubalaena japonica OR12G1
Steno bredanensis OR2A1ψ
Steno bredanensis OR2A2ψ
Steno bredanensis OR13K1ψ
Stenella coeruleoalba (AJ233789) OR4D1ψ
Delphinus delphis OR6A1ψ
Stenella coeruleoalba OR6A1
Delphinus delphis OR2AT1ψ
Orcinus orca OR2AT1ψ
Pseudorca crassidens OR2AT1ψ
Steno bredanensis OR2AT1ψ
Stenella coeruleoalba (AJ233791) OR2AT1ψ
Phocoena phocoena OR2AT1Aψ
Phocoena phocoena OR2AT1Bψ
Delphinus delphis OR11M1ψ
Stenella coeruleoalba (AJ233804) OR10P1ψ
Delphinus delphis OR1I1ψ
Stenella coeruleoalba OR1I1Aψ
Pseudorca crassidens OR1I1ψ
Steno bredanensis OR1I1ψ
Orcinus orca OR1I1ψ
Stenella coeruleoalba (AJ233790) OR1I1Bψ
Phocoena phocoena OR1I1ψ
Physeter macrocephalus OR1I1ψ
Eubalaena japonica OR1I1ψ
Phocoena phocoena OR1F1ψ
Stenella coeruleoalba (AJ233802) OR1F1ψ
Phocoena phocoena OR1F2ψ
Stenella coeruleoalba OR1AB1
Eubalaena japonica OR7E1
Orcinus orca OR7E1ψ
Delphinus delphis OR7G1ψ
Orcinus orca OR7D1ψ
Phocoena phocoena OR7D1ψ
Eubalaena japonica OR7C1
Physeter macrocephalus OR7C1ψ
Eubalaena japonica OR7A1
Delphinus delphis OR10J1ψ
Stenella coeruleoalba OR10J1Aψ
Stenella coeruleoalba OR10J1Bψ
Pseudorca crassidens OR10J1ψ
Steno bredanensis OR10J1ψ
Orcinus orca OR10J1ψ
Pseudorca crassidens OR10J3ψ
Stenella coeruleoalba OR10J3ψ
Physeter macrocephalus OR10J4ψ
Delphinus delphis OR10J2ψ
Pseudorca crassidens OR10J2ψ
Steno bredanensis OR10J2ψ
Stenella coeruleoalba OR10J2ψ
Phocoena phocoena OR10J2ψ
Orcinus orca OR10K1ψ
Delphinus delphis OR10K1
Stenella coeruleoalba OR10K1
Stenella coeruleoalba (AJ233792) OR10K1
Pseudorca crassidens OR10K1
Phocoena phocoena OR10K1
Physeter macrocephalus OR10K1 ψ
Physeter macrocephalus OR10K2Aψ
Physeter macrocephalus OR10K2Bψ
Eubalaena japonica OR10K2
Orcinus orca OR10K3ψ
Steno bredanensis OR10K3
Delphinus delphis OR10K3
Pseudorca crassidens OR10AB1ψ
Delphinus delphis OR10AB1ψ
Steno bredanensis OR10AB1ψ
Stenella coeruleoalba OR10AB1Aψ
Stenella coeruleoalba OR10AB1Bψ
Phocoena phocoena OR10AB1ψ
Orcinus orca OR10AB1Aψ
Orcinus orca OR10AB1Bψ
Physeter macrocephalus OR10AB1Aψ
Physeter macrocephalus OR10AB1Bψ
Physeter macrocephalus OR10AB1Cψ
Eubalaena japonica OR10AB1Aψ
Eubalaena japonica OR10AB1Bψ
Orcinus orca OR10AB2ψ
Delphinus delphis OR10AB2ψ
Delphinus delphis OR6M1ψ
Stenella coeruleoalba OR6M1ψ
Pseudorca crassidens OR6M1ψ
Orcinus orca OR6M1ψ
Steno bredanensis OR6M1Aψ
Steno bredanensis OR6M1Bψ
Phocoena phocoena OR6M1ψ
Physeter macrocephalus OR6M1ψ
Steno bredanensis OR6M2ψ
Stenella coeruleoalba OR6M2ψ
Pseudorca crassidens OR10A1
Steno bredanensis OR10A1
Stenella coeruleoalba OR10A1
Stenella coeruleoalba (AJ233793) OR10A1
Eubalaena japonica OR10AG1
Delphinus delphis OR5P1ψ
Physeter macrocephalus OR5P1ψ
Delphinus delphis OR5P2ψ
Steno bredanensis OR5P2
Steno bredanensis OR5BH1
Stenella coeruleoalba OR5BH1
Stenella coeruleoalba (AJ233803) OR5AW1ψ
Eubalaena japonica OR8H1
Eubalaena japonica OR5D1
Eubalaena japonica OR5D2
Eubalaena japonica OR8T1
Stenella coeruleoalba OR8T1
Delphinus delphis OR8B1ψ
Stenella coeruleoalba (AJ233801) OR8B1ψ
Stenella coeruleoalba (AJ233799) OR8B2ψ
Steno bredanensis OR9I1ψ
Stenella coeruleoalba (AJ233805) OR9Q1ψ
0.05 substitutions/site

FIGURE 1. Bayesian consensus phylogram of 127 cetacean olfactory receptor (OR) genes. Putative OR gene identification is listed to the right of each species name, with inferred pseudogenes in red lettering and denoted by ψ. Hypothesized functional genes are in blue. Gray circles at nodes indicate maximum parsimony (MP) and maximum likelihood (ML) bootstrap scores ≥70% and Bayesian posterior probability ≥0.95. Asterisks above nodes denote topological conflict with MP and/or ML trees. Support scores within orthologue groups are not shown.

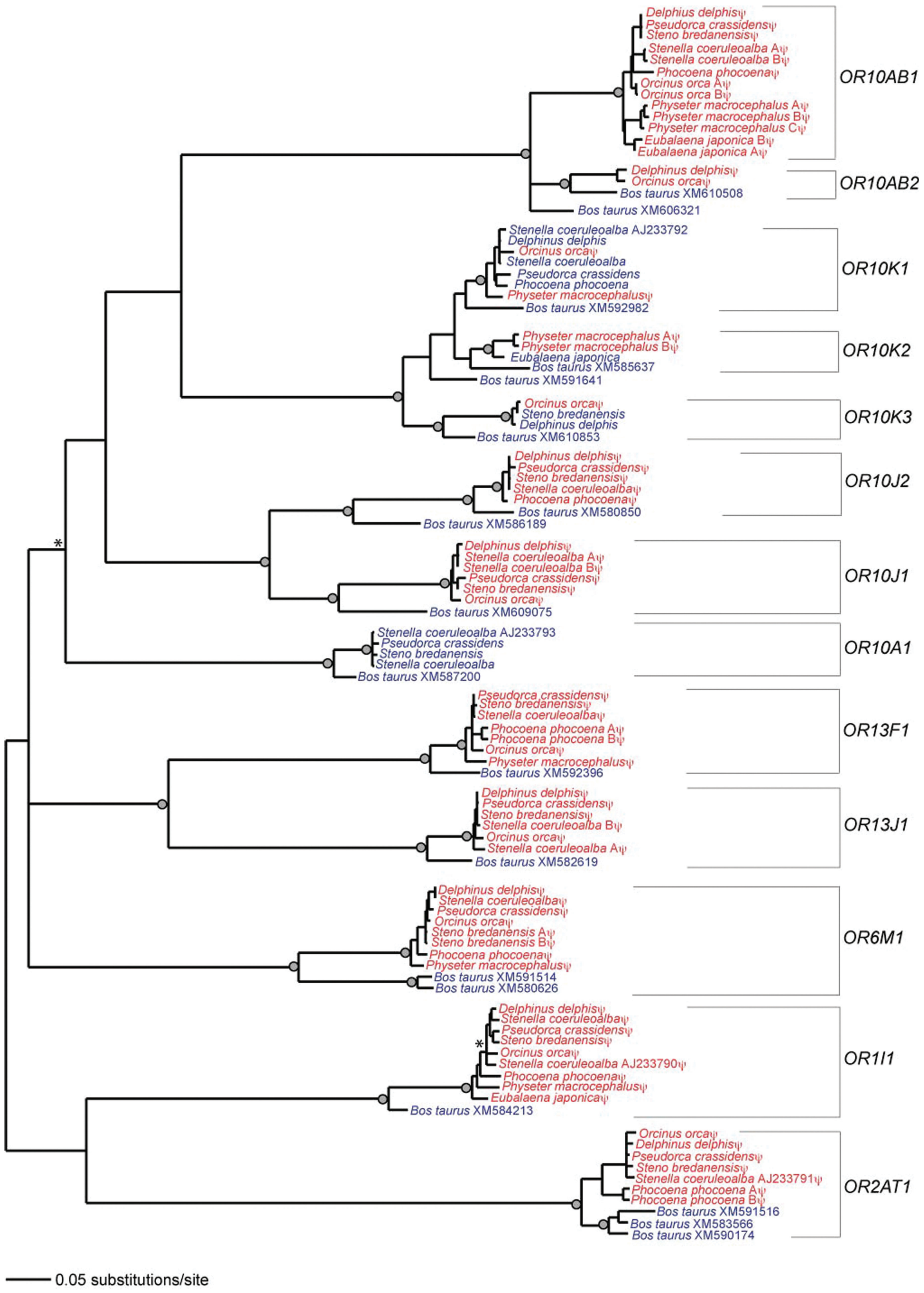

FIGURE 2. Bayesian consensus phylogram of the 13 best-represented cetacean OR orthologue groups and closely related *Bos* sequences. Putative identities of OR orthologue groups are shown at the right. Inferred pseudogenes are in red lettering and denoted by ψ. Hypothesized functional genes are in blue. Gray circles at nodes indicate MP and ML bootstrap scores ≥70% and Bayesian posterior probability ≥0.95. Asterisks above nodes denote topological conflict with MP and/or ML trees. Most support scores within orthologue groups are not shown.

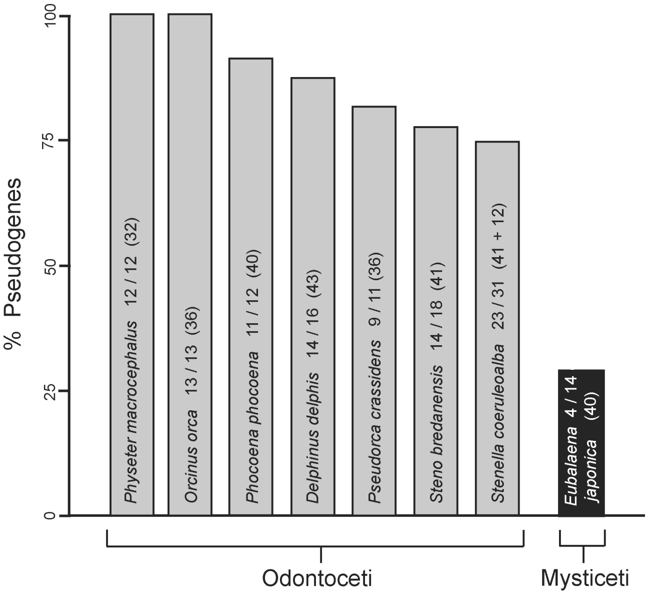

FIGURE 3. Percentage of pseudogenes to total OR genes sampled for each cetacean species. Number of pseudogenes over total number of genes sampled is shown in each bar followed by number of clones sequenced in parentheses. For *Stenella coeruleoalba,* the total number of clones includes 12 sequences from Freitag et al. (1998).

(Fig. 1). The inclusion of 12 *Stenella coeruleoalba* sequences from Freitag et al. (1998) increased the number of pseudogenes by 10 to a total of 100 out of 127 (∼78.7%).

Pseudogenes were not distributed evenly among species (Fig. 3). The data supported a major reduction in the number of functional OR genes in toothed whales compared to the baleen whale *Eubalaena*, corresponding to absence of the olfactory apparatus in mature odontocetes (Oelschläger, 1992). On average, odontocete OR genes consisted of ∼85.0% pseudogenes. At one extreme, all OR genes sampled from *Orcinus orca* (killer whale) and *Physeter macrocephalus* (sperm whale) were inferred pseudogenes. The sampled OR repertoire of *Stenella coeruleoalba* (striped dolphin) contained the lowest percentage of pseudogenes, ∼74.2% (Fig. 3). These data are consistent with OR pseudogene proportions estimated from two other odontocetes, *Kogia sima* (10/13 = 76.9%) and *Phocoenoides dalli* (7/9 = 77.7%; Kishida et al., 2007).

In contrast to the seven odontocetes, only 4 of 14 OR genes (∼28.6%) sampled from the mysticete, *Eubalaena japonica*, were inferred pseudogenes (Fig. 3). This percentage is lower than the OR pseudogene proportion documented in another mysticete, *Balaenoptera acutorostrata* (11/19 = 57.9%; Kishida et al., 2007), but both mysticete species are well below proportions for all odontocetes sampled thus far. The larger percentage of apparently functional OR genes in mysticetes, and especially *Eubalaena,* is consistent with retention of a reduced olfactory apparatus in mysticetes through maturity (Oelschläger, 1992). The pseudogene proportion estimated for *Eubalaena* is actually much lower than that of *Homo sapiens* (Rouquier et al., 2000; Gilad et al., 2004), although sample size is comparatively small (Fig. 3).

Overall, our genetic survey of Cetacea showed that pseudogenes were distributed throughout the OR superfamily tree (Figs. 1, 2). Phylogenetic analyses of individual orthologue groups suggested that various OR gene lineages were silenced at different periods in cetacean history (Fig. 4). In some cases, such as *OR6M1*, all cetaceans sampled for a particular orthologue shared a common ancestral frameshift indel (Fig. 4b); a similar pattern was observed in *OR10J1*, *OR10J2*, and *OR13J1*. In contrast, based on our data, pseudogenization was the result of independent mutations in different cetacean lineages for *OR1I1*, *OR10AB1*, and *OR13F1*; however, it cannot be ruled out that a single, common silencing event occurred outside the region sequenced. Given the current database, some silencing events occurred in multiple odontocete lineages in parallel, particularly between Delphinoidea (dolphins and porpoise) and Physeteridae (sperm whale; e.g., *OR1I1* and*OR10AB1*; Fig. 4). These results are consistent with fossil evidence for independent loss of olfaction in divergent lineages of odontocetes (Kellogg, 1928; Hoch, 2000; Geisler and Sanders, 2003), but larger samples of complete OR genes from multiple species are necessary to further corroborate this hypothesis. Nevertheless, it is evident that pseudogenization has occurred in multiple odontocete OR genes, resulting in mass silencing of a gradually dying gene superfamily.

OR Pseudogenes: Rates of Evolution

Several authors have noted a significant reduction in the rate of nucleotide substitution in cetacean DNA sequences relative to other placental mammals (Martin and Palumbi, 1993; Bininda-Emonds, 2007); thus, there is a need for nuclear loci with enough variation to resolve recently diverged or rapidly radiating lineages in the group. For this application, OR pseudogenes are promising phylogenetic markers due to their relatively high rate of divergence and in theory, a lack of distortion from selection (Li et al., 1981; Gojobori et al., 1982; Ophir and Graur, 1997). Pairwise comparisons (Table 4) of new and published sequences from two delphinids (*Orcinus orca* and *Delphinus delphis*) demonstrated that in these taxa, OR pseudogenes (∼1.69% divergence) show slightly more variation than the nuclear introns sampled here (∼1.68%) and have evolved at approximately three times the rate of the nuclear exons (∼0.62%). The

TABLE 4. General time-reversible (GTR) pairwise distances and average number of indels for *Orcinus orca* versus *Delphinus delphis* sequences. Results are shown separated by partition with number of genes and total aligned bases listed.

Partition	Number of genes	Total aligned bases	% GTR pairwise distance	Indels per 1000 bp
Olfactory receptors	9	4642	1.69	1.51
Nuclear introns	5	2486	1.68	0.80
Nuclear exons	7	4042	0.62	0.00
mt rDNA genes	2	950	2.06	2.11
mt D-loop	1	537	8.72	7.44
mt Protein-coding	2	1824	9.23	0.00

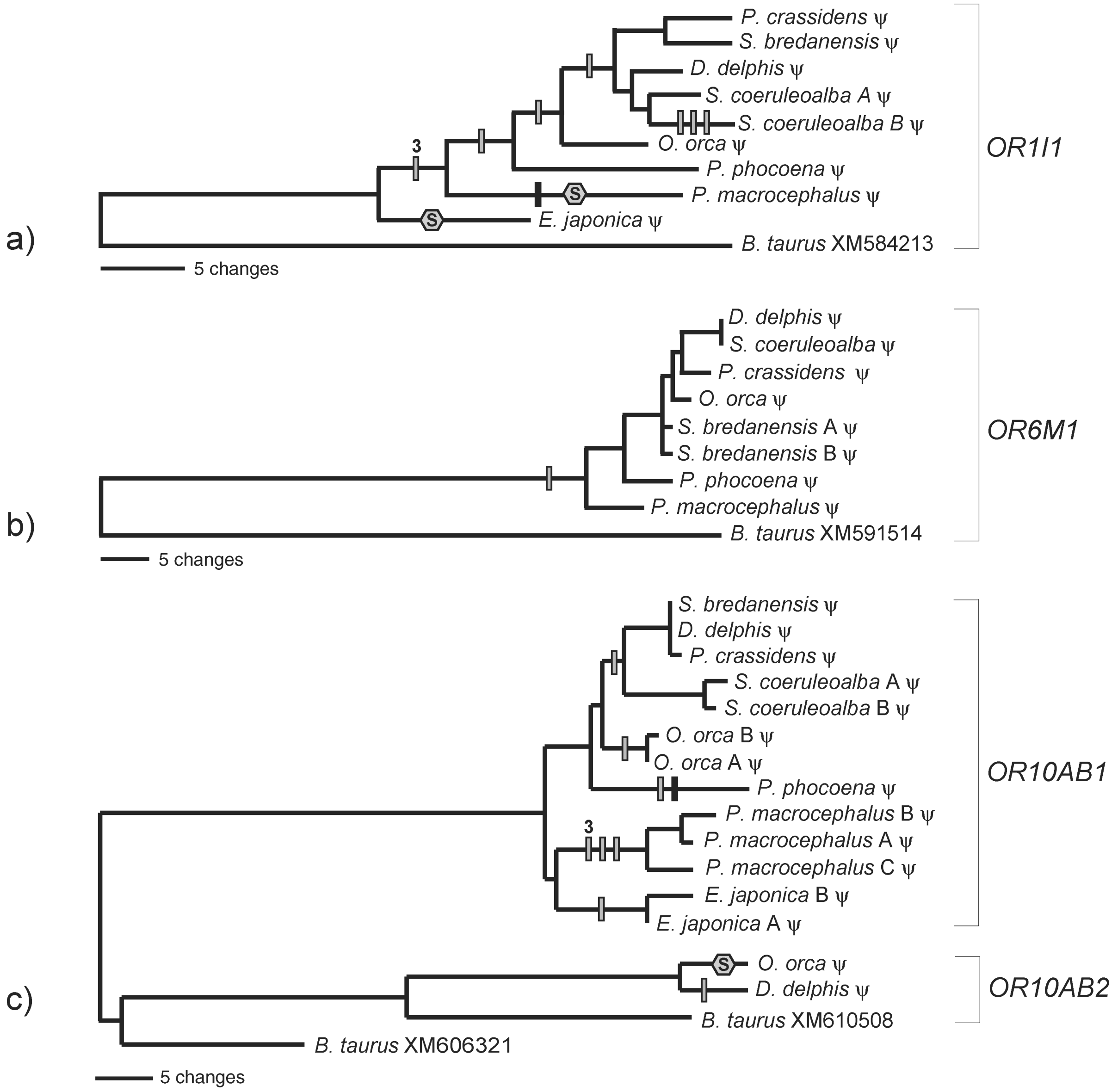

FIGURE 4. Maximum parsimony gene trees and indel mappings for four OR orthologue groups, (a) *OR1I1*, (b) *OR6M1*, (c) *OR10AB1* and *OR10AB2*. Indel events are shown as gray bars (deletions) and black bars (insertions). In-phase indels are identified by a "3" above bars. Nucleotide substitutions that created stop codons are symbolized by gray hexagons containing an "S."

OR pseudogenes were characterized by more indels (1.51/1000 bp) than either nuclear introns (0.80/1000 bp) or nuclear exons (0.00/1000 bp). mtDNA generally evolves at a much faster rate than nuDNA in mammals (e.g., Brown et al., 1982; Springer et al., 2001), and in our study, OR pseudogene divergence was predictably much less than in D-loop sequences and mitochondrial protein-coding genes (Table 4). Overall, the data for Delphinidae suggest that OR pseudogenes have evolved at a rapid rate in comparison to nuclear exons, at a slightly faster rate than nuclear introns, and slower than different classes of mtDNA.

Supermatrix Analyses of OR Genes and Independent DNA Data

Gene trees for individual OR orthologue groups illustrated differences in topology and degree of resolution (Fig. 5). Despite substantial missing data (Fig. 6d) and conflicts among genes (Figs. 4 and 5), the concatenated matrix of 11 OR orthologue groups resulted in a well-supported and fully resolved tree that was generally consistent with traditional classifications of Cetacea (Fig. 6a). The MP tree and the Bayesian consensus tree were topologically congruent (5755 characters; 109 parsimony informative; consistency index [CI] = 0.952; retention

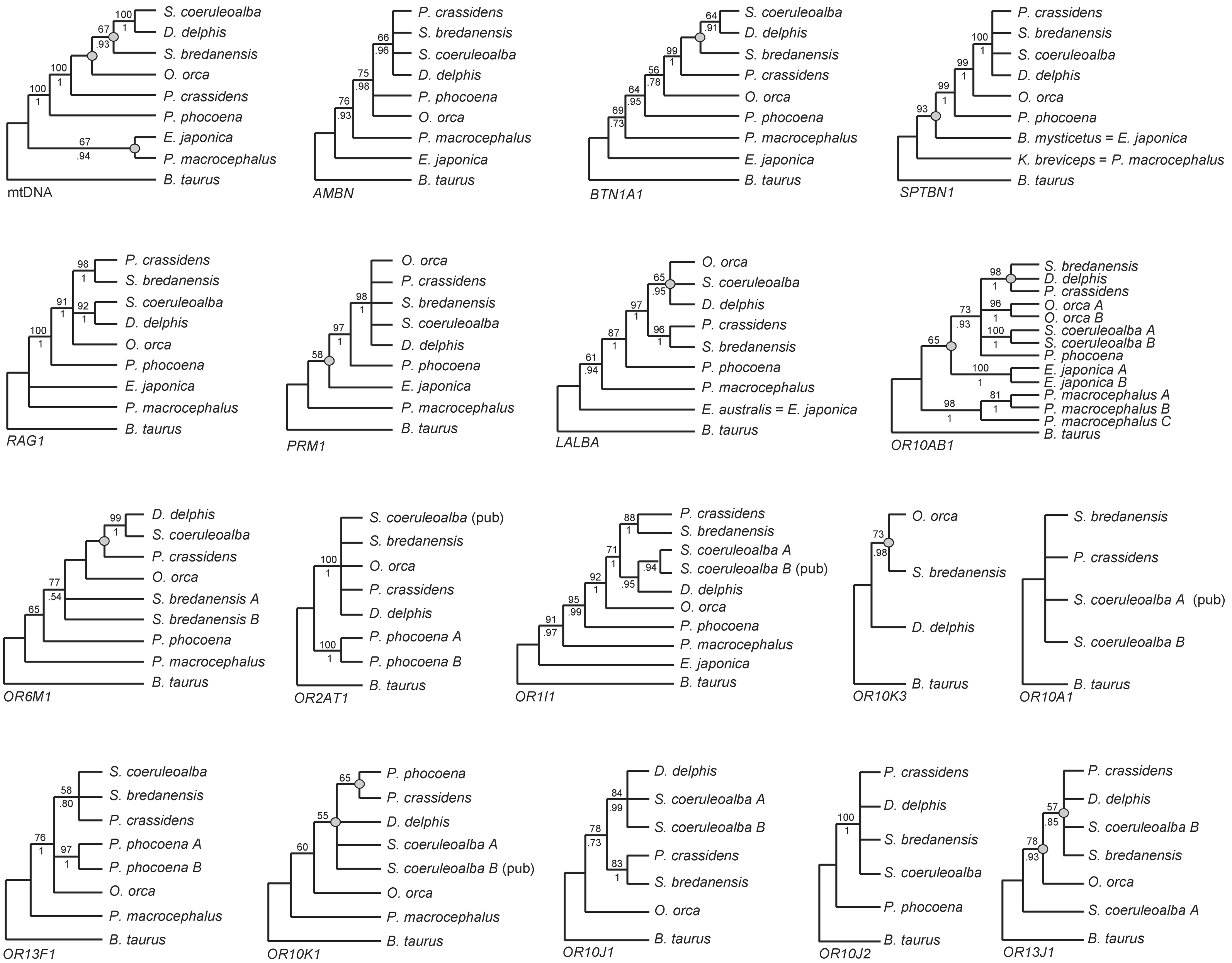

FIGURE 5. Maximum parsimony gene trees for mitochondrial (mt) DNA, single-copy nuclear genes, and OR genes. Numbers above branches are parsimony bootstrap support scores, and numbers below branches show Bayesian posterior probabilities. Gray circles identify nodes that conflict with the combined data Bayesian tree (Figure 6d).

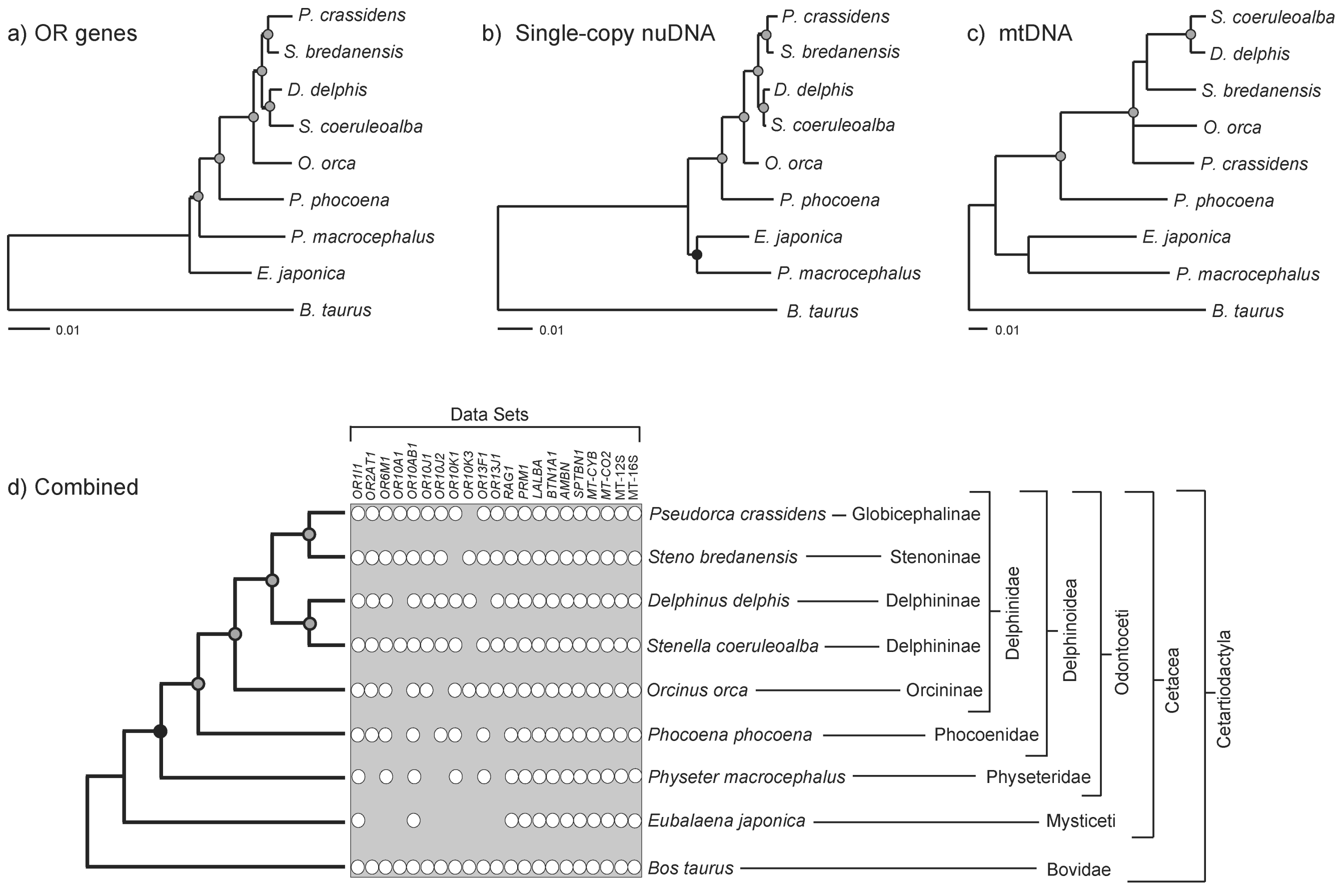

FIGURE 6. Bayesian consensus phylograms for (a) combined OR genes, (b) combined single-copy nuclear (nu) DNA, and (c) combined mtDNA. Scale is 0.01 subsitution per site. (d) Bayesian consensus tree for the supermatrix (OR genes + single-copy nuDNA + mtDNA genes). Traditionally recognized taxonomic groups are to the right of species names. White circles in the gray box indicate which genes were sampled for each taxon. In all trees (a–d), circles at nodes indicate level of support: Bayesian posterior probability ≥0.95 (black) or both MP bootstrap score ≥70% and Bayesian posterior probability ≥0.95 (gray). In analyses that included OR genes, support scores differed only slightly when *Stenella coeruleoalba* sequence B was replaced with sequence A for the *OR13J1* gene.

index [RI] = 0.693). All nodes were supported by parsimony bootstrap scores of ≥78% and Bayesian posterior probabilities of ≥0.99 (Fig. 6a). A separate parsimony analysis of sequence data alone (without gap characters included) recovered the same topology.

We compared the OR topology with two independent data sets: single-copy nuDNA and mtDNA (Fig. 6b, c). The Bayesian consensus topology for the single-copy nuDNA (Fig. 6b) agreed with that of the OR nuDNA, with one exception: *Eubalaena* and *Physeter* were more closely related in the tree based on single-copy genes, rendering Odontoceti paraphyletic. Instability at the base of crown-group Cetacea has been discussed in many previous analyses (e.g., Milinkovitch et al., 1994); however, odontocete paraphyly is strongly contradicted by multiple morphological synapomorphies and 12 SINE insertions (Geisler and Sanders, 2003; Nikaido et al., 2007). Parsimony analysis of the single-copy genes resulted in the same topology as the OR tree (Fig. 6a) and supported odontocete monophyly (3547 characters; 132 parsimony informative; CI = 0.935; RI = 0.784). Relationships within Delphinidae were identical in trees derived from the OR data (Fig. 6a), single-copy nuDNA (Fig. 6b), and the combination of these nuDNA data sets (not shown).

The topology and support values of the mtDNA tree (Fig. 6c) differed substantially from the nuDNA trees (Fig. 6a, b). Only three nodes had high support values, compared to six nodes for the combined OR data and five nodes for the single-copy nuDNA. Unlike both nuDNA trees, some relationships within Delphinidae were unresolved or very weakly supported. Our results were similar to published analyses of *MT-CYB* (LeDuc et al., 1999; May-Collado and Agnarsson, 2006); the addition of more mtDNA data did little to increase resolution or support among major lineages of Delphinidae, and the data showed higher variability and homoplasy (2803 characters; 456 parsimony informative; CI = 0.722; RI = 0.508) relative to nuDNA. One explanation for the weak performance of mtDNA is a tendency toward saturation of rapidly evolving sites on long terminal branches separated by short internodes, a pattern evinced by other groups that have undergone rapid radiations in the past (Kraus and Miyamoto, 1991; Allard et al., 1992).

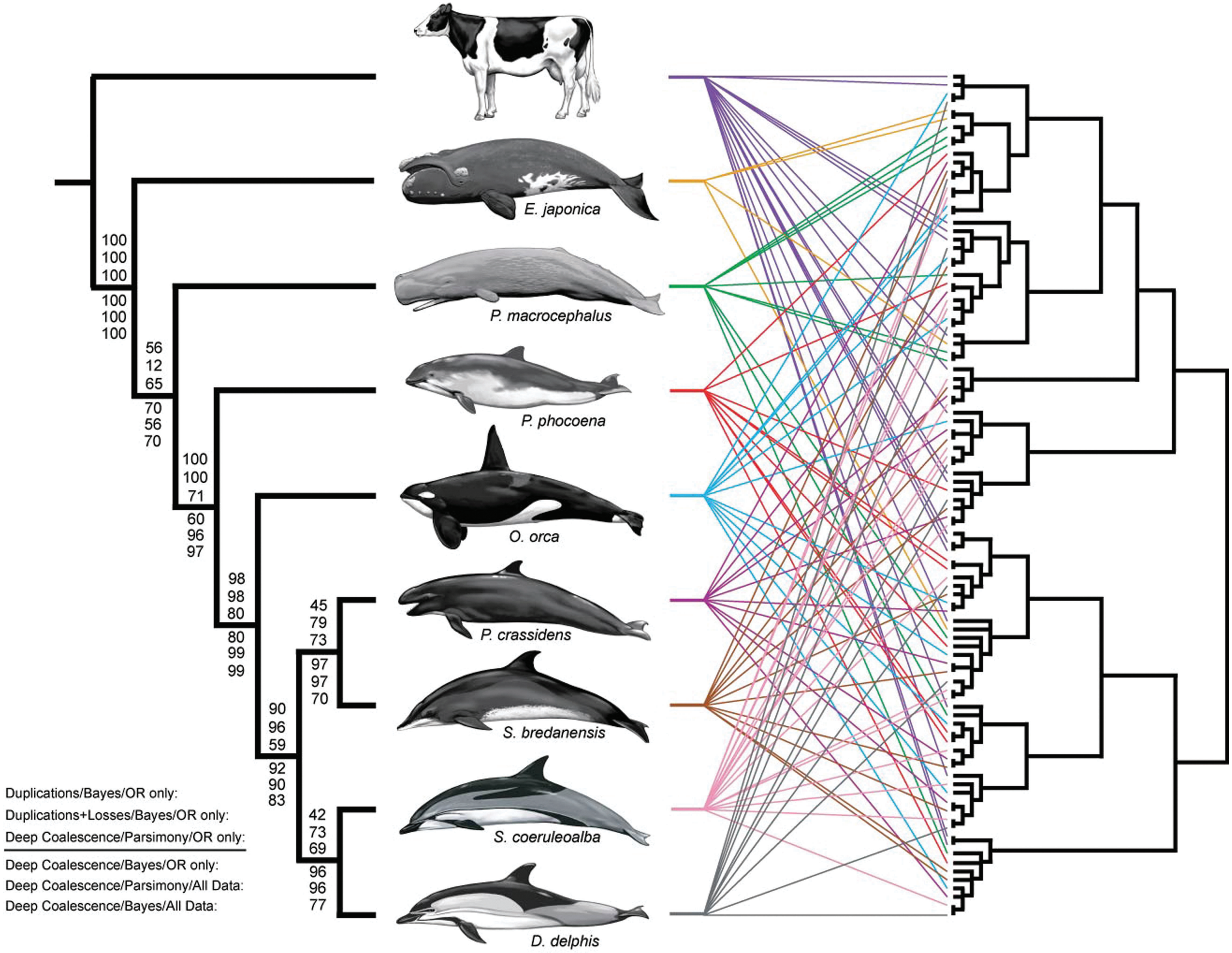
100
100
100
100
100
100
56
12
65
70
56
70
100
100
71
60
96
97
98
98
80
80
99
99
90
96
59
92
90
83
45
79
73
97
97
70
42
73
69
96
96
77
E. japonica
P. macrocephalus
P. phocoena
O. orca
P. crassidens
S. bredanensis
S. coeruleoalba
D. delphis
Duplications/Bayes/OR only:
Duplications+Losses/Bayes/OR only:
Deep Coalescence/Parsimony/OR only:
Deep Coalescence/Bayes/OR only:
Deep Coalescence/Parsimony/All Data:
Deep Coalescence/Bayes/All Data:

FIGURE 7. Species tree (left) resulting from gene-tree reconciliation (GR) analyses. A representative gene tree of the cetacean OR superfamily is shown on the right with colored lines linking genes with their respective species. Gene-tree bootstrapping support for each GR analysis is shown above and below each branch. Key in bottom left corner corresponds with support values at internodes (optimization criterion/method of analysis of gene trees/data sets). In the analysis of duplications only for the OR data, two bootstrap scores within Delphinidae were below 50% (42% and 45%), but support for conflicting clades was low (≤21%).

The Bayesian combined analysis of all data sets resulted in a fully resolved tree (Bayesian posterior probabilities of 1.0; Fig. 6d). All robustly supported nodes in the MP analysis (12105 characters; 697 parsimony-informative; CI = 0.838; RI = 0.572) agreed with the combined Bayesian supermatrix analysis, which was identical to the OR topology (Fig. 6a). Relationships within Delphinidae were uniform across analyses of OR genes, single-copy nuclear genes, all nuDNA, and all DNA combined.

In the past, *Steno* (rough-toothed dolphin, Stenoninae) has been placed within Delphininae in many traditional taxonomies (Kasuya, 1973; de Muizon, 1988), and at least one example exists of hybridization between *Steno* and a delphinine (*Tursiops truncatus*) in captivity (Dohl et al., 1974). In our study, the combined mtDNA and the nuclear gene *BTN1A1* did place *Steno* with *Stenella coeruleoalba* and *Delphinus delphis* (Delphininae) in a weakly supported clade (Figs. 5 and 6c). However, all of our combined analyses that included nuclear loci instead supported a close relationship between *Steno* and *Pseudorca crassidens* (Globicephalinae), a result recently corroborated by Caballero et al. (2008). This novel clade was primarily supported by two OR genes (*OR1I1* and *OR10J1*) and two single-copy nuclear genes (*LALBA* and *RAG1;* Fig. 5). In all combined analyses here that included nuDNA, representatives of Delphininae, Globicephalinae, and Stenoninae grouped robustly to the exclusion of *Orcinus orca* (killer whale, Orcininae; Fig. 6), consistent with some molecular analyses of Harlin-Cognato and Honeycutt (2006).

GR Analyses of OR Genes and Independent DNA Data

Overall, GR analyses of the OR data (Fig. 7) and independent loci were highly congruent with the supermatrix results (Fig. 6). The inferred species trees generally supported the monophyly of Delphinidae and three delphinid subclades: Delphininae (*Delphinus delphis* +*Stenella coeruleoalba*), a grouping of *Steno bredanensis* (Stenoninae) + *Pseudorca crassidens* (Globicephalinae), and a grouping of *S. bredanensis* + *P. crassidens* + Delphininae to the exclusion of *Orcinus orca* (Orcininae) as in the OR and combined supermatrix trees (Fig. 6a and d). At a higher taxonomic level, Delphinoidea (Delphinidae + the phocoenid, *Phocoena phocoena*) and Cetacea were monophyletic. A *Eubalaena* and *Physeter* clade was supported by the GR analysis that minimized gene duplications plus losses (gene-tree bootstrap = 88%), but all other GR searches grouped *Physeter* with the other members of Odontoceti.

The mammalian OR gene superfamily is expansive, and as such is a challenging test for both supermatrix and GR methods of phylogeny reconstruction. Despite substantial missing data, rate accelerations in non-functional gene lineages, and conflicts among loci, both supermatrix and GR analyses of ORs recovered the same species tree in seven of eight analyses (Figs. 6, 7). Instead of viewing supermatrix and GR supertrees as incompatible, these can be interpreted as complementary estimates of phylogeny that reciprocally illuminate both character support and partition support for a combined data set (Bininda-Emonds, 2004b). Future studies will be necessary to determine whether the strong congruence between methods recorded in our analysis is the norm.

In conclusion, our results supported the hypothesis that a diverse, functional OR subgenome (Fig. 1) has become severely reduced in modern odontocete whales, concomitant with the loss of anatomical structures associated with olfaction. OR genes produced a well-supported phylogenetic hypothesis, especially among lineages of oceanic dolphins that have been difficult to place in previous studies (Figs. 6, 7); the addition of more delphinid species will serve as a future test of these relationships. Supermatrix and GR analyses of the OR database consistently recovered the same topology. Furthermore, nucleotide substitutions and indels in OR genes showed a low degree of homoplasy in comparison to mtDNA and overwhelming congruence with independent single-copy nuDNA. These patterns suggest that gene duplications (Figs. 1, 2), ancestral lineage sorting, rate heterogeneity, and introgression have not obscured phylogenetic signal in this case study. Although the use of OR pseudogenes in phylogenetics has yet to be tested in other groups, such pseudogenes are common in all vertebrate genomes studied to date (Rouquier et al., 2000; Zhang and Firestein, 2002; Gilad et al., 2003, 2004; Olender et al., 2004; Niimura and Nei; 2005). Defunct OR genes represent many rapidly diverging nuclear loci for future use in phylogenetic reconstruction and may be especially effective for resolving relationships among recently diverged vertebrates that have undergone rapid diversifications.

Acknowledgments

Funding was provided by NSF grants (DEB-0212572, DEB-0640313, DEB-0213171) and the University of California Riverside Academic Senate. We thank the Marine Mammal Center (Sausalito, CA), P. Morin, K. Robertson, and R. LeDuc of Southwest Fisheries Science Center, NOAA (La Jolla, CA), and M. Milinkovitch for providing DNA samples. C. Hayashi, N. Ayoub, J. Garb, O. Bininda-Emonds, J. Cotton, and an anonymous reviewer provided helpful comments on various drafts of this manuscript. Discussion with T. Rowe on anatomical correlates of pseudogene formation helped to inspire this study.

References

Allard, M. W., M. M. Miyamoto, L. Jarecki, F. Kraus, and M. R. Tennant. 1992. DNA systematics and evolution of the artiodactyl family Bovidae. Proc. Natl. Acad. Sci. USA 89:3972–3976.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Barnes, L. G. 2002. Delphinoids, evolution of the modern families. Pages 314–316 *in* Encyclopedia of marine mammals (W. F. Perrin, B. Würsig, J. G. M. Thewissen, eds.). Academic Press, San Diego.

Bininda-Emonds, O. R. P. 2004a. The evolution of supertrees. Trends Ecol. Evol. 19:315–322.

Bininda-Emonds, O. R. P. 2004b. Trees versus characters and the supertree/supermatrix "paradox." Syst. Biol. 53:356–359.

Bininda-Emonds, O. R. P. 2007. Fast genes and slow clades: Comparative rates of molecular evolution in mammals. Evol. Bioinformatics 3:59–85.

Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J. Mol. Evol. 18:225–239.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

Caballero, S., J. Jackson, A. A. Mignucci-Giannoni, H. Barrios-Garrido, S. Beltrán-Pedreros, M. G. Montiel-Villalobos, K. M. Robertson, and C. S. Baker. 2008. Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. Mol. Phylogenet. Evol. 46:252–268.

Carlini, D. B., K. S. Reece, and J. E. Graves. 2000. Actin gene family evolution and the phylogeny of coleoid cephalopods (Mollusca: Cephalopoda). Mol. Biol. Evol. 17:1353–1370.

Cave, A. J. E. 1988. Note on olfactory activity in mysticetes. J. Zool. Lond. 214:307–311.

Cotton, J., and R. D. M. Page. 2002. Going nuclear: Gene family evolution and vertebrate phylogeny reconciled. Proc. R Soc. B 269:1555–1561.

Cotton, J., and R. D. M. Page. 2003. Gene tree parsimony vs. uninode coding in phylogenetic reconstruction. Mol. Phylogenet. Evol. 29:298–308.

de Jong, W. W., and L. Ryden. 1981. Causes of more frequent deletions than insertions in mutations and protein evolution. Nature 290:157–159.

Deméré, T. A., M. R. McGowen, A. Berta, and J. Gatesy. 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. Syst. Biol. 57:15–37.

de Muizon, C. 1988. Les relations phylogénétiques des Delphinida (Cetacea, Mammalia). Ann. Paléontol. 74:159–227.

de Queiroz, A., and J. Gatesy. 2007. The supermatrix approach to systematics. Trends Ecol. Evol. 22:34–41.

Dohl, T. P., K. S. Norris, and I. Kang. 1974. A porpoise hybrid: *Tursiops* X *Steno*. J. Mammal. 11:207–221.

Durand, D., B. V. Halldorsson, and B. Vernot. 2006. A hybrid micro-macroevolutionary approach to gene tree reconstruction. J. Comput. Biol. 13:320–335.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Freitag, J., G. Ludwig, I. Andreini, P. Rössler, and H. Breer. 1998. Olfactory receptors in aquatic and terrestrial vertebrates. J. Comp. Physiol. A 183:635–650.

Gatesy, J., C. Hayashi, M. A. Cronin, and P. Arctander. 1996. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. Mol. Biol. Evol. 13:954–963.

Geisler, J. H., and A. E. Sanders. 2003. Morphological evidence for the phylogeny of Cetacea. J. Mammal. Evol. 10:23–129.

Gilad, Y., O. Man, S. Pääbo, D. Lancet. 2003. Human specific loss of olfactory receptor genes. Proc. Natl. Acad. Sci. USA 100:3324–3327.

Gilad, Y., V. Wiebe, M. Przeworski, D. Lancet, and S. Pääbo. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. PLoS Biol. 2:120–125.

Glusman, G., A. Bahar, D. Sharon, Y. Pilpel, J. White, and D. Lancet. 2000. The olfactory receptor gene superfamily: Data mining, classification, and nomenclature. Mammal. Genome 11:1016–1023.

Glusman, G., I. Yanai, I. Rubin, and D. Lancet. 2001. The complete human olfactory subgenome. Genome Res. 11:685–702.

Gojobori, T., W.-H. Li, and D. Graur. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. 18:360–369.

Goodman, M., J. Czelusniak, G. W. Moore, A. E. Romero-Herrera, and G. Matsuda. 1979. Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. 28:132–163.

Guindon, S., and O. Gascuel. 2003. A simple and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Hamilton, H, S. Caballero, A. G. Collins, and R. L. Brownell Jr. 2001. Evolution of river dolphins. Proc. R. Soc. B 268:549–556.

Harlin-Cognato, A. D., and R. L. Honeycutt. 2006. Multi-locus phylogeny of dolphins in the subfamily Lissodelphininae: Character synergy improves phylogenetic resolution. BMC Evol. Biol. 6:87.

Hoch, E. 2000. Olfaction in whales: Evidence from a young odontocete of the Late Oligocene North Sea. Historical Biol. 14:67–89.

Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11:152–158.

Irwin, D. M., T. D. Kocher, and A. C. Wilson. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. 32:128–144.

Kasuya, T. 1973. Systematic consideration of recent toothed whales based on morphology of tympano-periotic bone. Sci. Rep. Whales Res. Inst. 25:1–103.

Kellogg, R. A. 1928. The history of whales—Their adaptation to life in the water. Q. Rev. Biol. 3:29–76, 174–208.

Kishida, T., S. Kubota, Y. Shirayama, and H. Fukami. 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.

Kraus, F., and M. M. Miyamoto. 1991. Cladogenesis among the pecoran ruminants: Evidence from mitochondrial DNA sequences. Syst. Zool. 40:117–130.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

LeDuc, R. G., W. F. Perrin, and A. E. Dizon. 1999. Phylogenetic relationships among delphinid cetaceans based on full cytochrome *b* sequences. Mar. Mammal Sci. 15:619–648.

Li, W.-H., T. Gojobori, and M. Nei. 1981. Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239.

Machado, C., and J. Hey. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. Proc. R. Soc. Lond. B 270:1193–1202.

Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Martin, A. P., and T. M. Burg. 2002. Perils of paralogy: Using HSP70 genes for inferring organismal phylogenies. Syst. Biol. 51:570–587.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA 90:4087–4091.

Mathews, S., and M. J. Donoghue. 2000. Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. Int. J. Plant Sci. 161:541–555.

Matthee, C. A., J. D. Burzlaff, J. F. Taylor, and S. K. Davis. 2001. Mining the mammalian genome for artiodactyl systematics. Syst. Biol. 50:367–390.

Matthee, C. A., and S. K. Davis. 2001. Molecular insights into the evolution of the family Bovidae: A nuclear DNA perspective. Mol. Biol. Evol. 18:1220–1230.

May-Collado, L., and I. Agnarsson. 2006. Cytochrome *b* and Bayesian inference of whale phylogeny. Mol. Phylogenet. Evol. 38:344–354.

Milinkovitch, M. C., M. Bérubé, and P. J. Palsbøll. 1998. Cetaceans are highly derived artiodactyls. Pages 113–132 *in* The emergence of whales (J. G. M. Thewissen, ed.). Plenum Press, New York.

Milinkovitch, M. C., A. Meyer, and J. R. Powell. 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. Mol. Biol. Evol. 11:939–948.

Miyamoto, M. M., and W. M. Fitch. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

Mombaerts, P., 2004. Genes and ligands for odorant, vomeronasal and taste receptors. Nat. Rev. Neurosci. 5:263–278.

Montgelard, C., F. M. Catzeflis, and E. Douzery. 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome *b* and 12S rRNA mitochondrial sequences. Mol. Biol. Evol. 14:550–559.

Müller, K. 2005. SeqState—Primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinformatics 4:65–69.

Niimura, Y., and M. Nei. 2003. Evolution of olfactory receptor genes in the human genome. Proc. Natl. Acad. Sci. USA 100:12235–12240.

Niimura, Y., and M. Nei. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc. Natl. Acad. Sci. USA 102:6039–6044.

Nikaido, M., O. Piskurek, and N. Okada. 2007. Toothed whale monophyly reassessed by SINE insertion analysis: The absence of lineage sorting effects suggests a small population of a common ancestral species. Mol. Phylogenet. Evol. 43:216–224.

Nylander, J. A. A. 2004. MrModelTest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Oelschläger, H. A. 1992. Development of the olfactory and terminalis systems in whales and dolphins. Pages 141–147 *in* Chemical signals in vertebrates VI (R. L. Doty and D. Müller-Schwarze, eds.). Plenum Press, New York.

Olender, T., T. Fuchs, C. Linhart, R. Shamir, M. Adams, F. Kalush, M. Khen, and D. Lancet. 2004. The canine olfactory subgenome. Genomics 83:361–372.

Ophir, R., and D. Graur. 1997. Patterns and rates of indel evolution in processed pseudogenes from humans and murids. Gene 205:191–202.

Page, R. D. M. 1998. GeneTree: Comparing gene and species phylogenies using reconciled trees. Bioinformatics 14:819–820.

Page, R. D. M. 2000. Extracting species trees from complex gene trees: Reconciled trees and vertebrate phylogeny. Mol. Phylogenet. Evol. 14:89–106.

Page, R. D. M., and M. A. Charleston. 1997. From gene to organismal phylogeny: Reconciled trees and the gene tree/species tree problem. Mol. Phylogenet. Evol. 7:231–240.

Posada, D., and T. R. Buckley 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793–808.

Posada, D., and K. A. Crandall. 1998. ModelTest: Testing the model of DNA substitution. Bioinformatics 9:817–818.

Queralt, R., R. Adroer, R. Oliva, R. J. Winkfein, J. D. Retief, and G. H. Dixon. 1995. Evolution of protamine P1 genes in mammals. J. Mol. Evol. 40:601–607.

Rambaut A., and A. J. Drummond. 2007. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Trace

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Ronquist, F., J. P. Huelsenbeck, and P. van der Mark. 2005. MrBayes 3.1 manual. Draft 26 May 2005. Distributed by the authors, http://mrbayes.csit.fsu.edu/manual.php.

Rouquier, S., A. Blancher, and D. Giorgi. 2000. The olfactory receptor gene repertoire in primates and mouse: Evidence for reduction of the functional fraction in primates. Proc. Natl. Acad. Sci. USA 97:2870–2874.

Simmons, M. P., C. D. Bailey, and K. C. Nixon. 2000. Phylogeny reconstruction using duplicate genes. Mol. Biol. Evol. 17:469–473.

Simmons, M. P., and J. V. Freudenstein. 2002. Uninode coding vs gene tree parsimony for phylogenetic reconstruction using duplicate genes. Mol. Phylogenet. Evol. 23:481–498.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49:369–381.

Slowinski, J. B., A. Knight, and A. P. Rooney. 1997. Inferring species trees from gene trees: A phylogenetic analysis of the Elapidae (Serpentes) based on the amino acid sequences of venom proteins. Mol. Phylogenet. Evol. 8:349–362.

Slowinski, J. B., and R. D. M. Page. 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48:814–825.

Springer, M. S., R. W. DeBry, C. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, and M. J. Stanhope. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. 18:132–143.

Steppan, S. J., R. M. Adkins, P. Q. Spinks, C. Hale. 2005. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Mol. Phylogenet. Evol. 37:370–388.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J. D., D. G. Higgins, and T. G. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position–specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Whinnett, A., and N. I. Mundy. 2003. Isolation of novel olfactory receptor genes in marmosets (*Callithrix*): Insights into pseudogene formation and evidence for functional degeneracy in non-human primates. Gene 304:87–96.

Zhang, X., and S. Firestein. 2002. The olfactory receptor gene superfamily of the mouse. Nat. Neurosci. 5:124–133.

First submitted 7 August 2007; reviews returned 1 October 2007; final acceptance 22 April 2008
Associate Editor: Olaf Bininda-Emonds