Molecular Phylogenetics and Evolution 79 (2014) 249–257

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Classification of the cormorants of the world

Martyn Kennedy, Hamish G. Spencer *

Allan Wilson Centre, Department of Zoology, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand

ARTICLE INFO

Article history:
Received 9 April 2014
Revised 16 June 2014
Accepted 18 June 2014
Available online 30 June 2014

Keywords:
Cormorants
Shags
Phalacrocorax
Phalacrocoracidae
Phylogeny
Taxonomy

ABSTRACT

Relationships among the 40 or so extant species of cormorants (family Phalacrocoracidae) have been obscured by their morphological similarities, many of which have recently been shown to be the result of convergent evolution. Previous attempts to derive an evolutionarily justifiable classification for this group of birds using osteological and behavioral data have been hampered by these similarities. We present a well-resolved evolutionary tree for some 40 cormorant taxa based on the results of extensive genetic work that produced over 8000 bases of mitochondrial and nuclear DNA sequence. This tree implies a novel classification for the cormorants, which reflects their evolutionary history and can be implemented using some 7 genera. Some of the relationships among the species are well-known but many are previously unrecognized. Nevertheless, much of the classification makes sense in terms of biogeography.

1. Introduction

Cormorants are a speciose group of large water birds characterized by long bills, largely monochromatic (white and dark black/gray/brown) plumage and webbed ("totipalmate") feet, which comprise the family Phalacrocoracidae. With an almost worldwide distribution, they are important predators of small fish in both marine and freshwater environments. The cormorants have a long independent history, with >40 million years separating them from the darters (*Anhinga*) (Gibb et al., 2013; see also Worthy, 2012). Relationships among the 40-or-so extant species have been obscured by their apparent morphological conservatism: it is easy to recognize a cormorant, but specific identification is often difficult (especially if geographical location is unknown). We have recently shown that these morphological similarities are due to an extraordinary degree of concerted evolution, to the point where phylogenetic trees estimated from these characters have strong statistical support for clades of species that are, on the basis on genetic data, unrelated (Holland et al., 2010).

One consequence of this convergent selection is that the true phylogenetic relationships among the species of cormorant have been obscure. This obscurity has, in turn, impacted cormorant taxonomy. Many treatments (e.g. Marchant and Higgins, 1990; American Ornithologists' Union, 2013; Remsen et al., 2013) avoid giving any indication of relationships among the species by using the single genus *Phalacrocorax*. In the past 40 years, however, two schemes for an evolutionarily based taxonomy have been proposed. Based on an informal analysis of morphological and behavioral features, van Tets (1976) argued that the family Phalacrocoracidae consisted of two groups, cormorants and shags, which he recognized at the generic level as *Phalacrocorax* and *Leucocarbo*, respectively. Within each of these genera he accepted several subgenera (see Table 1). Cormorants were largely associated with freshwater and coastal environments and shags with the open sea. Siegel-Causey (1988) used a cladistic analysis of osteological characters to derive a remarkably similar classification. Although he held that cormorants and shags were differentiable at the higher (subfamily) level, most of van Tets's (1976) subgenera corresponded to Siegel-Causey's genera. The exception was van Tets's subgenus *Leucocarbo*, from which Siegel-Causey removed the marine cormorants (cormorants that, because of their marine habitat, looked like shags), as well as several other groups that he recognized as separate genera (see Table 1).

Siegel-Causey's classification implied some rather implausible biogeography, however, and subsequent genetic data have revealed that many of his genera are not monophyletic (Kennedy et al., 2000, 2001, 2009; Holland et al., 2010). Even the division of the Phalacrocoracidae into cormorants and shags, while influential and adopted by various monographs (e.g. Johnsgard, 1993) and ornithological checklists (e.g. Gill et al., 2010), appears, on the basis of genetics, to be illusory (Kennedy et al., 2000; Holland et al., 2010).

Here we present the results of a phylogenetic analysis of over 8000 base pairs of mitochondrial and nuclear DNA sequences for almost all the extant species of phalacrocoracids. From our

* Corresponding author.
E-mail address: hamish.spencer@otago.ac.nz (H.G. Spencer).

http://dx.doi.org/10.1016/j.ympev.2014.06.020
1055-7903/

Table 1
Summary and comparison of van Tets's (1976) and Siegel-Causey's (1988) classifications (after Johnsgard (1993)).

Van Tets (1976)	Siegel-Causey (1988)
Genus *Phalacrocorax*: Cormorants	Subfamily Phalacrocoracinae
Subgenus (*Phalacrocorax*): Macro-Cormorants	Genus *Phalacrocorax*: Macro-Cormorants
Subgenus (*Hypoleucos*): Meso-Cormorants	Genus *Hypoleucos*: Meso-Cormorants
Subgenus (*Microcarbo*): Micro-Cormorants	Genus *Microcarbo*: Micro-Cormorants
[part of *Leucocarbo* s. str]	Genus *Compsohalieus*: Marine Cormorants
Genus *Leucocarbo*: Shags	Subfamily Leucocarbinae: Shags
Subgenus (*Leucocarbo*): King Shags	Genus *Leucocarbo*: Guano Shags
Subgenus (*Leucocarbo*): King Shags	Genus *Notocarbo*: Blue-eyed Shags
Subgenus (*Leucocarbo*): King Shags	Genus *Euleucocarbo*: New Zealand Blue-eyed Shags
Subgenus (*Leucocarbo*): King Shags	Genus *Notocarbo*: Campbell Island Shag
Subgenus (*Stictocarbo*): Cliff Shags	Genus *Stictocarbo*: Cliff Shags

analysis, we derive an evolutionary taxonomy for the cormorants of the world, implemented through seven monophyletic genera. Some of the relationships among the species are well-known but many are previously unrecognized. Nevertheless, much of the classification makes sense in terms of biogeography.

2. Materials and methods

Tissue, blood or feathers were obtained from a number of sources (see Table 2). Given the relationships found in Hackett et al. (2008), samples from a pelican, gannet, booby and darter were selected for use as outgroups. Total genomic DNA was extracted from each of the samples using a phenol/chloroform extraction, a 5% Chelex 100 solution or the DNeasy Tissue Kit (Qiagen) (e.g. see Kennedy and Spencer, 2004; Spencer et al., 2006; Kennedy et al., 2013). Negative controls were included with each extraction. Following extraction, the DNA was amplified for five mitochondrial genes (the 12S ribosomal RNA gene [12S], the overlapping ATPase-8 and -6 genes [ATPase], the NADH dehydrogenase subunit 2 gene [ND2] and the cytochrome oxidase subunit I gene [COI]) and five nuclear genes (β-fibrinogen intron 7 [FIB7], Parkinson disease 7 [PARK7], interferon regulatory factor 2 [IRF2], crystallin alpha A [CRYAA], and rap guanine nucleotide exchange factor 1 [RAPGEF1]). The polymerase chain reaction (PCR) was used to amplify these regions with 45 °C annealing for ATPase-8 and -6 and the barcoding region of COI, 50 °C annealing for the other COI fragment, ND2 and FIB7, 55 °C annealing for 12S and IRF2, 56 °C annealing for PARK7, 62 °C annealing for RAPGEF1, and 63 °C annealing for CRYAA (slightly lowered annealing temperatures were occasionally used depending on the marker and the template). For 12S, ATPase and part of COI we used the primers and followed the procedures described in Kennedy and Spencer (2004), for example, using the primers COIf and COIa (see Palumbi, 1996) for COI. For the other COI fragment (the barcoding region) we used either the primer pair BirdF1 and BirdR1 or (more commonly) FalcoFA and VertR1 from Kerr et al. (2007). This barcoding fragment overlapped with the fragment produced by the COIf and COIa primer pair, and they could subsequently be combined (or were sometimes amplified as a single piece using the FalcoFA and COIa primer combination). For ND2 the primer pair Av5199tMetF and Av6314tTrpR were used (see Kennedy et al., 2013). For FIB7 the primers FIB-BI7U and FIB-BI7L (Prychitko and Moore, 1997) were used. For the other nuclear markers the primers used all came from Kimball et al. (2009): PARK7 (PARK.2F and PARK.3R), IRF2 (IRF2.2F and IRF2.3R), CRYAA (CRY.1F and CRY.2R), and RAPGEF1 (RAP.18F and RAP.19R).

The PCR conditions were an initial denaturation step of 94 °C (3 min), followed by 40 cycles of 94 °C (30 s), variable annealing temperatures (as described above) for 45 s to 1 min, and 72 °C (1 min) and a final extension phase at 72 °C for 4 min. Negative controls were included with each PCR reaction. The PCR products were either left unpurified or were purified using the PureLink PCR purification kit (Invitrogen), the High Pure PCR purification kit (Roche), or the Ultra-Sep Gel extraction kit (Omega), and then sequenced on an automated sequencer using the PCR primers (sometimes using internal sequencing primers for 12S, see Kennedy and Spencer, 2004).

The mitochondrial sequences were aligned by eye following the procedure outlined in Kennedy et al. (2000), whereas the nuclear sequences were initially aligned using ClustalX 2.0 (Larkin et al., 2007) using the default settings before being finally aligned by eye. Any positions where the alignment was uncertain were excluded to avoid mistaken homology, while any unique insertions were also excluded (e.g. in the one extreme case the darter had a 668 base insertion in PARK7). The sequences are available from GenBank (for the Accession Nos. see Table S1) and the data matrix and resultant phylogenetic tree from TreeBASE (www.treebase.org). Phylogenetic analyses were performed with MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) for Markov-chain Monte-Carlo Bayesian analysis and posterior probabilities, PAUP* version 4b10 (Swofford, 2002) for maximum parsimony (MP) bootstrap searches (Felsenstein, 1985) and PhyML 3.0 (Guindon et al., 2010) for maximum likelihood (ML) bootstrap searches. We used the partition-homogeneity test (Swofford, 2002) to investigate whether the different gene sequences contain similar signals and could thus be analyzed as a single data set (with the MulTrees option unchecked). For visualization purposes the pelican was defined as the outgroup taxon (see Gibb et al., 2013; Hackett et al., 2008).

The models of nucleotide substitution for the Bayesian analysis were selected using the Akaike Information Criterion of Modeltest 3.7 (Posada and Crandall, 1998). The models selected for each gene region were: TIM + I + G for 12S and ATPase (6st + I + G), GTR + I + G for ND2 (6st + I + G), K81uf + I + G for COI (6st + I + G), TVM + G for FIB7 (6st + G), GTR for PARK7 (6st), GTR + G for IRF2 (6st + G), and HKY + G for CRYAA and RAPGEF1 (2st + G).

Bayesian analysis was performed using MrBayes v3.1.2 with the maximum likelihood model employing either 6 or 2 substitution types ("nst = 6" or "nst = 2") for each partition (see above). For 12S, ATPase, ND2 and COI rate variation across sites was modeled using a gamma distribution, with a proportion of the sites being invariant ("rates = invgamma"). For FIB7, IRF2, CRYAA and RAPGEF1 rate variation across sites was modeled using a gamma distribution, with none the sites being invariant ("rates = gamma"). For PARK7 the model selected had no rate variation across sites, and none of the sites were invariant ("rates = equal"). Trees were estimated for all of the partitions combined. For the combined