An holistic ecological analysis of the diet of Cory's shearwaters using prey morphological characters and DNA barcoding

Hany Alonso[1,2,4,*], José Pedro Granadeiro[3], Silke Waap[1,5], José Xavier[2], William O. C. Symondson[5], Jaime A. Ramos[2] and Paulo Catry[1]

[1]Eco-Ethology Research Unit, ISPA, Rua Jardim do Tabaco 34, 1149-041 Lisboa, Portugal.

[2]Institute of Marine Research (IMAR/CMA), Department of Life Sciences, University of Coimbra, 3004-517 Coimbra, Portugal.

[3]Centro de Estudos do Ambiente e do Mar (CESAM)/Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal.

[4]Museu Nacional História Natural e da Ciência, Universidade de Lisboa, Rua da Escola Politécnica 58, 1250-102 Lisboa, Portugal.

[5]Cardiff School of Biosciences, Sir Martin Evans Building, Cardiff University, Museum Avenue, Cardiff CF10 3AX, UK

*Email: hany_alonzo@hotmail.com

Key-words: *Calonectris diomedea*, non-breeders, pre-breeding, prey identification, Selvagens, sexual segregation

Corresponding author: Hany Alonso

Running title: DNA barcoding and diet of a seabird

Abstract

Knowledge of the dietary choices and trophic niches of organisms is the key to understanding their roles in ecosystems. In seabird diet studies, prey identification is a difficult challenge, often yielding results with technique-specific biases. Additionally, sampling efforts are often not extensive enough to reveal intra-populational variation. Immature animals, which may constitute up to 50% of a population, may occupy a significantly different trophic niche to more-experienced birds, but this remains largely unexplored. We investigated the diet of Cory´s shearwater (*Calonectris diomedea*) from Selvagem Grande, an island located off the northwest African coast, collecting a total of 698 regurgitate samples over three consecutive breeding seasons. The diet was assessed using two complementary approaches for prey identification: conventional morphological analysis (using fish vertebrae, otoliths and cephalopod beaks) and DNA barcoding of the 16S rRNA mitochondrial gene, in cases where a positive identification could not be retrieved. Species assignments employed BLAST and distance based methods, as well as direct optimization of the tree length based on unaligned sequences in POY. This method resulted in robust tree estimates and species assignments, showing its potential for DNA barcoding of stomach contents using hypervariable markers such as the 16S. The molecular approach increased taxonomic resolution and revealed an additional 17 taxa. Diet differed significantly according to breeding status, sex, breeding phase (pre-laying and chick-rearing) and year. Such direct evidence of trophic segregation within the same population has rarely been shown in seabirds and highlights the importance of including such variables in ecosystem-based management approaches.

Introduction

Dietary studies are essential building blocks of the science of ecology. Only with the support of dietary studies can we properly assess the position of species in food webs, their role in energy flow within ecosystems, the importance of feeding resources for demographic regulation and the impact of predation on populations and communities. Despite this, the trophic niche of numerous ecologically important species, such as top marine predators, remains poorly understood (e.g., Naito *et al.* 2013). Two primary problems in previous research have been difficulties with prey identification and failure to sample relevant population segments (Barrett *et al.* 2007; Bowen *et al.* 2013) that may potentially display niche differentiations (Polis 1984). Amongst seabirds, which are major pelagic consumers, much effort has gone into sampling the diet of breeding birds (often only at the chick stage) while virtually nothing is known about non-breeders (Barrett *et al.* 2007). This gap is particularly relevant when one considers that non-breeders (mostly immature individuals) may represent >50% of the fully grown individuals in a population.

Trophic niche differentiation between immature and adult reproducing individuals is to be expected in species where growth is protracted and niche is strongly influenced by body size (e.g., Lucifora *et al.* 2009). However, in other taxa, particularly in birds, size varies little amongst fledged individuals. Nevertheless, even for a relatively invariable body size, we may expect differences linked to, for example, (a) age-related improvements in foraging competence (Kitowski 2003; Daunt *et al.* 2007), (b) the competitive exclusion of subdominant (generally younger) individuals by more dominant conspecifics (Goss-Custard *et al.* 1982), or (c) differential spatial distribution arising from the need of reproducing individuals to regularly attend breeding sites. Despite these expectations, we currently know very little about

whether there are ontogenetic changes of trophic niches in birds, or their possible causes and consequences.

Molecular techniques, such as DNA barcoding (Kochzius *et al.* 2010; Zhang & Hanner 2012), are revolutionising dietary studies and are now being extensively applied in dietary analyses of vertebrate and invertebrate carnivores and herbivores (reviewed in Symondson 2002; Pompanon *et al.* 2012). Prey species can be identified even from highly degraded tissue (as found in faeces and regurgitates), using PCR. Most of these studies have identified prey species from homogenised meta-samples (guts or faeces), with quantitative estimates of species consumed derived from sequences obtained for each identified prey using either a cloning and sequencing technique or Next Generation Sequencing. Nevertheless, differences among prey species in the mitochondrial copy numbers per cell, as well as in the binding efficiency of the primers (Symondson 2002; Pompanon *et al.* 2012), may lead to substantial biases. One way to overcome this problem is to use a combined approach, using morphological analyses for quantitative estimates of prey (hard parts recovered from guts or faeces) plus augmentation of species identification using DNA barcoding of tissues (Barnett *et al.* 2010; Dunn *et al.* 2010). Applying this approach to pelagic top-predators has the potential to enhance understanding of trophic dynamics and, as such, marine conservation and ecosystem-based management.

Birds are amongst the best studied animal classes, yet few studies have used molecular techniques to improve our understanding of their trophic ecology (e.g., Deagle *et al.* 2007; Jedlicka *et al.* 2013). Recently, molecular approaches has been used to investigate the dietary habits of seabirds, but those few studies have analysed faeces only (Deagle *et al.* 2007; Bowser *et al.* 2013; Jarman *et al.* 2013; but see Jarman *et al.* 2002), which implies that quantification of identified prey remained relatively crude (Deagle *et al.* 2010). The first aim

of the present paper is to develop the technique and illustrate the tremendous potential of using DNA barcoding combined with morphological tools to provide an unusually refined picture of the diet of birds (in this case, of a pelagic seabird).

Our study model is the Cory's shearwater (*Calonectris diomedea borealis*), an oceanic predator of the Northeast Atlantic, which breeds on several islands and islets from the Azores and Berlengas archipelagos in the north to the Canary archipelago in the south (Thibault *et al.* 1997). The feeding ecology of Cory's shearwaters has been studied at several colonies (Granadeiro *et al.* 1998; Paiva *et al.* 2010; Xavier *et al.* 2011; Neves *et al.* 2012), but little is known about their diet in the southern area of their breeding range (but see den Hartog & Clarke 1996; Paiva *et al.* 2010 for studies with limited sampling effort). More importantly, Cory's shearwaters are long-lived birds that only start reproducing at a mean age of 9 years and frequently skip breeding seasons, even after their first reproduction (Mougin *et al.* 1997). As such, a large proportion of Cory's shearwater populations is comprised of non-breeders, but their trophic ecology has never been investigated. We also have a poor understanding of male-female differences in the ecology of this dimorphic species (Navarro *et al.* 2009; Ramos *et al.* 2009) and most studies carried out found no evidence of spatial (Navarro *et al.* 2009) or foraging niche sexual segregation (Navarro *et al.* 2007; Ramos *et al.* 2009). Hence, the second broad objective of this paper is to characterise the diet of Cory's shearwaters in the southern part of its breeding range and assess within-population sources of variation, with a particular interest in differentiation between breeders and non-breeders, indicative of ontogenetic shifts in the trophic niche of this seabird.

Methods

Study area and species

Fieldwork was conducted in the Selvagem Grande island (30° 09' N, 15° 52' W), where ca. 30,000 Cory's shearwater breeding pairs nest (Granadeiro *et al.* 2006). This sub-tropical oceanic island is located ca. 350 km from up-welling enriched shelf areas off the African coast. Cory's shearwaters are long-distance migrants returning in early March from their wintering sites in the south Atlantic (Thibault *et al.* 1997). During the extended pre-laying period, birds re-occupy their nest cavities, protecting them from prospecting birds and eventually finding a mate. Egg-laying takes place at the end of May with the chicks hatching at the end of July. The chick-rearing period lasts approximately 97 days, until early November, when chicks fledge (Thibault *et al.* 1997).

Diet sampling and analysis

Sampling was conducted in the pre-laying period of 2010 (11 to 20 April) and during the chick-rearing periods of 2008, 2009 and 2010 (28 July to 2 October). Shearwaters returning from the sea were captured by hand when entering the nest or preparing to feed their chick. Non-breeding birds were also sampled in the chick-rearing period of 2009. In the incubation period, when birds are more sensitive to disturbance and are more likely to have empty stomachs, only non-breeders were sampled (15 to 25 June of 2010). Non-breeding Cory's shearwaters tend to stay outside the nest cavities and to be very vocal and socially interactive, and are, therefore, easily selected for sampling.

Birds were sexed based on their distinctive vocalisations (Thibault *et al.* 1997) or using a discriminant function based on bill measurements, which has a 98.8% success rate (Granadeiro 1993).

Prey samples were collected from the birds using the water off-loading technique (Wilson 1984). By selecting different areas each day and marking the birds with wax markers, we guaranteed that birds and nests were only sampled once. Fresh prey items found in food samples were identified using specialised guides (Quéro *et al.* 2003) and stored in 70% ethanol. Digested fish were quantified and identified to the lowest possible taxon from vertebrae and other hard remains (otoliths, dentaries and scales), using our own reference collection and published guides (Tuset *et al.* 2008). Cephalopods were identified from their beaks and quantified based on the number of mantles, other fresh remains (tentacles, flesh) and fresh beaks (upper and lower beaks were counted).

A large number of fish from the genus *Scomber* were identified as *Scomber colias* (41.1%, N = 538). None was identified as *Scomber scombrus* and it was only possible to identify the remaining individuals to genus level (*Scomber*). Given this result, we pooled *Scomber colias* and *Scomber* sp. in all further analyses. We calculated frequencies of occurrence (FO): the number of samples with a given prey type, expressed as a percentage of the total number of samples and numerical frequencies (NF): the number of individuals of a given taxon, as a percentage of the total number of prey items.

Given their small size (less than 3 mm), most unidentified crustaceans and insects found in the samples were unlikely to be their direct prey, and were probably part of the diet of fish captured by shearwaters (secondary predation). Considering their parasitic habits, crustaceans from the family Isopoda were also probably captured along with fish prey. None of these prey were included when calculating the numerical importance of prey. The

exception were three larger crustaceans (more than 30 mm, Decapoda) that were considered to be part of the shearwater diet.

Genetic analysis

A total of 83 muscle samples (27 cephalopods and 56 fish), either unidentified through conventional diet analysis (45 samples) or only identified to higher taxonomic levels (*Trachurus* sp. and Exocoetidae) (38 samples), were examined using DNA barcoding (16s rRNA). Although the cytochrome c oxidase subunit I (COI) has gathered wide consensus as a genetic marker for species discrimination of unknown taxa (Hebert *et al.* 2003), the 16S barcode provided a higher sequence database coverage within the range of prey identified in Cory´s shearwater diet. For example, in teleosts, all genera within the family Exocoetidae were covered for the 16S, but only three for the COI. A search on squid "Teuthida" in the GenBank database retrieved 359 matches against 305 (after excluding the family Loliginidae, which is by far the best represented in Genbank). Therefore, the 16S was more informative in the context of this study.

We collected pieces of tissue from prey associated with hard structures (e.g., vertebrae) that could not be identified morphologically and used these for DNA barcoding. To extract prey DNA, individual prey tissue was washed with ddH_2O to remove adherent ethanol. As in other barcoding studies, that identified prey remains in stomach contents (e.g., Barnett *et al.* 2010), we chose where possible the inner parts of the tissue, since tissue from complex meta-samples may be contaminated with DNA of other prey. The DNA was extracted using the DNeasy Blood and Tissue Extraction Kit (Qiagen) following the protocol for purification of total DNA from animal tissues. Individual prey DNA from regurgitates was amplified using the universal primers of Palumbi (1996): 16ar, 5'-

CGCCTGTTTATCAAAAACAT-3' and 16br, 5'-CCGGTCTGAACTCAGATCACGT-3´, with an expected amplicon length of ca 550-620bp.

Polymerase Chain Reactions (PCR) were performed with the Multiplex PCR Kit (Qiagen) using the following PCR reagent mixtures: 10µl of Multiplex PCR Master Mix (1X), 0.25µM of each primer, 0.1mM of BSA, 3.6 µl ddH2O, 2.4µl (~50 – 100 ng/µl) of template DNA in a total volume of 20 µl. Thermal cycling conditions were as follow: 95ºC for 15min; 35 cycles of 94ºC for 30s, 52ºC for 90s, 72ºC for 90s, and a final extension at 72ºC for 10min. PCR products were cleaned using ExonucleaseI and Antarctic Alkaline phosphatase enzymes (New England, Biolabs) and sequenced using the EZ-seq services of Macrogen, Inc (Amsterdam, Netherlands).

Molecular identification of prey using BLAST

Sequences were compared with those in GenBank using the BLAST algorithm (Altschul *et al.* 1990). Each taxonomic assignment was based on the percentage of similarity with the reference sequences in GenBank. Species were directly assigned when the query sequence produced an identical match to the reference sequence (100% of identity). For BLAST matches higher than 99.0%, species were assigned when the query sequence matched monotypic genera or when the distribution range of potential con-specifics was outside our study area, but only if no other species was retrieved with this value. Inter-specific divergences in teleosts are > 2% (i.e., Kochzius *et al*. 2010, Zhang & Hanner 2012) and in cephalopods 1.3-12.7% (Dai *et al.* 2012). Therefore, the above criteria were expected to produce robust species identifications.

Molecular identification of prey using phylogenetic analysis

Specimens were assigned using phylogenetic inferences in cases where the percentage of similarity between the query sequence and the reference sequence was lower than 99.0%. Two methods were used: (a) distance based Neighbour-Joining (NJ) trees and (b) direct optimisation (DO) of the tree length.

NJ trees were constructed in Mega 6 (Tamura *et al.* 2013) using the Kimura-2-parameter model of evolution (Kimura 1980). The nodal support was obtained using a 1000 bootstrap replicates. NJ trees were estimated from eight different data sets of aligned sequences, each corresponding to the families that produced the nearest match with the query sequences. Sequences for which no positive identifications were obtained in BLAST were included and aligned with all available representative genera of that family using Clustal W (Thompson *et al.* 1994) as implemented in BioEdit (Hall 1999).

DO analysis were performed in POY v 5.0.0 (Varón *et al.* 2010). This program infers the tree directly from unaligned sequences and overcomes, therefore, potential uncertainties in sequence alignment (e.g., the hypervariable 16S mtDNA, where different numbers of indels between sequences can significantly impact tree estimates). To generate the POY tree we used the reference sequence that produced the nearest match in BLAST, including all other congener reference sequences of the same family. The tree estimated in POY did not include cephalopods since only two taxa could not be identified using BLAST (only represented in NJ trees). Sequences were trimmed to produce the exact same sequence terminals (374-396 base pairs), since sequences that are absent in the terminals can account for erroneous indel event counts in POY (De Laet 2010).

Because POY uses empirical gap cost criteria to optimise the tree length, we first performed sensitivity estimates under five different affine gap costs regimes: (2,1,1), (2,1,2),

(2,1,3), (2,1,5), (2,1,7) (substitution cost, gap extension, gap opening). Trees based on parsimony were constructed using 100 initial trees generated by random addition sequences using Subtree Pruning and Regrafting (SPR) and Tree Bisection and Reconnection (TBR) branch swapping. The tree producing the most congruent topology with what is known of the evolutionary relationships of teleost fish was chosen as the "optimum" tree. Nodal support was calculated using a 1000 bootstrap replicates with alternate SPR and TBR swapping.

Assignments of families and genera were obtained using a strict criterion based on how query sequences clustered in the NJ and DO trees (Wilson *et al.* 2011). According to this criterion, a taxon (family or genera) is identified if the query sequence nests within a clade that comprises members of that taxon.

As some of our query sequences that produced 100% matches in BLAST showed high similarities with other congeners (>98%) (families Carangidae and Exocoetidae), those were also included for phylogenetic analysis to validate species assignments. Species were identified if the query sequences clustered monophyleticaly with the taxon that produced an identical match in BLAST and with no other congeners. Unidentified vertebrae based on morphological analysis but otherwise positively identified using DNA barcoding were later used to identify those species and quantify their occurrence in all samples.

Statistical analysis

We initially checked for overall differences in the diet between sexes and among years using permutational multivariate analysis of variance based on distance matrices, implemented using the package "vegan" (Oksanen *et al.* 2011) running in R software (R Development Core Team 2010). The method undertakes a partitioning of the sums of squares of a multivariate data set, using semi-metric and metric distance matrices to produce a

256 "pseudo-F value". We tested for the effects of sex and year on the frequency of occurrence of
257 all prey items (with frequencies larger than 5% in one breeding phase). Whenever these tests
258 provided significant results, we further explored the effects of these factors (and their
259 interaction) on the occurrence of each prey using binomial GLMs (Generalized Linear
260 Models), with a logit link function. The statistical significance of each factor was tested
261 through log-likelihood ratio tests of increasingly simpler nested models, based on chi-squared
262 distributions.
263
264 **Results**
265
266 A total of 698 regurgitates were collected from adult Cory's shearwaters. From these
267 samples, a total of 2018 prey items were collected, 76.6% of which were successfully
268 identified to the species or genus level.
269 Morphological inspection of the 2018 prey items retrieved 40 different prey types, but
270 only 23 of those could be identified to species or genera. The use of DNA barcoding on
271 morphologically unidentifiable specimens increased the prey list to 17 new taxa (12 species, 3
272 genera and 2 families).
273
274 **Prey discrimination**
275 DNA barcoding of the16S ribosomal RNA gene produced longer fragments in teleost
276 fish than in cephalopods varying approximately between 550-600 and 460-500 base pairs,
277 respectively. DNA sequences were submitted to GenBank (Table S1).
278 BLAST comparisons in GenBank allowed for positive identification of 35% of the
279 sequences to the species level, while phylogenetic inferences successfully discriminated

another 46% to the genus level. From these sequences, 14 (10 species, 2 genera and 2 families) matched taxa that have never been identified in the diet of Cory´s shearwaters using morphological characters (Table 1). We also confirm the presence of the neon flying-squid (*Ommastrephes bartramii*) in the diet of these birds, where the beaks of small specimens were difficult to distinguish from those of the European flying-squid (*Todarodes sagittatus*).

It is noteworthy that the values of similarity between species and genera varied substantially, depending on the families and prey groups analysed. While most teleost families and cephalopods showed sequence homologies lower than 98% between conspecifics and congeners (within the reported divergences of vertebrate species), members of the families Exocoetidae and Carangidae presented very high homologies even between genera (ca. 99%). Therefore, identifications in both families were only obtained based on phylogenetic inferences. Regardless of the method employed for estimating trees (NJ or DO) the terminal topologies between the different trees were highly congruent (Fig. 1, Figs. S1, S2). Congeners clustered in highly supported monophyletic groups, with the exception of some members of the family Myctophidae and the genus *Cheilopogon*, that were paraphyletic and polyphyletic, respectively. Query sequences clustered, generally, with the reference sequences that produced the highest sequence homology in BLAST. Moreover, DO inferences resulted in a highly resolved tree at both internal and terminal nodes with an "optimal" tree obtained using the following settings: cost regime of substitutions = 2, indels = 1 and gap opening = 3. A total of seven major clades with high bootstrap support (85-100) were obtained, with each representing a different family of teleost fish.

Based on phylogenetic assignments using strict and liberal criteria we were also able to increase the taxonomic resolution of morphologically unidentified Exocoetidae and *Trachurus* specimens, identifying two species of *Cheilopogon* (*C. melanurus* and *C.*

304 *pinnatibarbatus*) and the species *Trachurus picturatus*. Furthermore, morphologically
305 unidentified members of the family Exocoetidae presented seven distinct Molecular
306 Operational Taxonomical Units (MOTUs), revealing a high diversity among these prey items.
307 Sequences of morphologically unidentified *Trachurus* specimens presented two distinct
308 MOTUs, where most sequences clustered separately from *T. picturatus* and the reference
309 sequences. The congruence in tree topologies as well as the taxonomic resolution obtained
310 suggests that genetic variability within the 16S rRNA gene is sufficient to discriminate
311 between species.
312
313 **Diet composition**
314 During the chick-rearing period, the diet of Cory's shearwaters was essentially
315 composed of fish (FO range = 88.7 to 91.1%) and cephalopods (FO range = 27.2 to 46.8%).
316 Chub mackerel (*Scomber colias*/sp.) was the most common prey (FO range = 35.6 to 51.2%)
317 (Table 2). Pilot-fish (*Naucrates ductor*; FO range = 13.3 to 16.1%), sardine (*Sardina*
318 *pilchardus*; FO range = 7.8 to 20.2%) and flying-fish (Exocoetidae; FO range = 7.4 to 14.4%)
319 also occurred frequently. Among flying-fish, two genera were found, namely *Exocoetus* (FO
320 range = 1.2 to 10%) and *Cheilopogon* (FO range = 1.1 to 3.2%). Subsequently, four species
321 were identified: the tropical two-wing flying-fish (*Exocoetus volitans*), the bandwing flying-
322 fish (*Cheilopogon exsiliens*), the Atlantic flying-fish (*C. melanurus*) and Bennett's flying-fish
323 (*C. pinnatibarbatus*). The diet of Cory's shearwaters was diverse, being composed of at least
324 33 fish species from 20 different families (Table 2). Unidentified fish were found in 13.3 to
325 19.7% of the samples mostly because they were too digested or lacked identifiable hard
326 remains.

The most common cephalopods were the neon flying-squid (*Ommastrephes bartramii*; FO range = 6.4 to 13.7%) and *Histioteuthis arcturi* (FO range = 1.1 to 5.3%). In addition, ten other cephalopod species from nine different families were present in the diet of Cory's shearwaters (see Table 2).

Crustaceans (FO range = 4.4 to 14.4%), mostly from the order Isopoda (FO range = 1.1 to 11.1%) and insects, from the family Halobatidae (FO range = 0 to 0.8%), were also present in the diet samples (Table 2). Fishery hooks were also found in three diet samples (FO = 0.5%).

Diet of non-breeders

During the chick-rearing period of 2009, the diet of non-breeders and breeders differed significantly (pseudo-$F_{1,240} = 3.63$, $p = 0.04$; Fig. 3). Non-breeders preyed heavily on cephalopods (FO = 63.2%), compared to breeders of the same year (FO = 35.1%), and consumed less chub mackerel (Fig. 3). During this period, neon flying-squid (FO = 36.8%), chub mackerel (FO = 31.6%), pilot-fish (FO = 21.1%) and horse/blue mackerel (FO = 10.5%) were the most frequent prey item of non-breeders (Fig. 3). In the incubation period, non-breeders also consumed less fish (FO = 12.1%) while cephalopods were much more frequent in their diet (FO = 93.9%), particularly neon flying-squid (FO = 45.5%).

Sex and inter-annual variations in diet

During the chick-rearing period, we found significant dietary differences between sexes (pseudo-$F_{1,513} = 10.63$ $p < 0.001$; Table 3). Females delivered significantly more chub mackerel to their chicks than males. In contrast, males provided the chicks with more sardines (Table 3). We also found significant inter-annual variations in the diet of shearwaters

(pseudo-$F_{2,513}$ = 11.52, $p < 0.001$), which were particularly noticeable in 2010, when the consumption of sardine and cephalopods was higher (Table 2), in comparison to previous years.

Seasonal variations in diet

The diet of shearwaters was substantially different among periods of the same year (pseudo-$F_{1,200}$ = 17.55, $p < 0.001$). In the pre-laying period of 2010, trumpet fish (*Macroramphosus scolopax*; FO = 36.7%, NF = 83.3%) and horse/blue jack mackerel (*Trachurus* sp.; FO = 40.0%, NF = 9.7%) dominated the diet of Cory's shearwaters (Fig. 2, Table 2). In this period, cephalopods were found to occur less frequently (FO = 13.3%, NF = 1.3%), than during the chick-rearing period of the same year (Fig. 2, Table 2). Other fish, particularly sardine and pilot-fish were frequent during the chick-rearing period, but completely absent from the diet during the pre-laying stage (Fig. 2, Table 2).

Discussion

This study of the diet of Cory's shearwaters from Selvagem Grande provides detailed information on the feeding ecology of this species. We used morphological and DNA barcoding methods in a complementary way to characterise and quantify the dietary composition of a pelagic seabird, showing the advantages of combining both techniques in diet studies of marine predators. The large number of samples collected during the provisioning period revealed dietary differences between birds of different breeding status and between sexes, something rarely shown to occur in seabirds.

The use of DNA barcoding in prey identification

DNA barcoding greatly improved our knowledge of Cory´s shearwater prey range, identifying species that would be overlooked in an analysis based solely on morphological traits. This was the case for small mesopelagic prey (such as myctophids), but also of some epipelagic and bathypelagic species that tend to be underestimated in morphological analyses due to a lack of representation in reference collections. Moreover, DNA barcoding proved to be effective in the identification of juvenile cephalopods, such as the neon flying-squid, a dominant prey in the diet of Cory's shearwater. Indeed, the identification of cephalopods from their beaks is challenging, particularly for small individuals, as many diagnostic characters only develop later in life.

In generalist predators such as seabirds, prey species can only be identified if a comprehensive database of reference sequences across different prey groups exists (Hebert *et al.* 2003). Mitochondrial 16S rRNA sequences are the most widely used marker in marine systems and provided the most extensive database of sequences relevant to the potential prey of Cory´s shearwaters. We found that inter- and intra-specific variability of the 16S barcode was effective for accurate prey species discriminations in teleosts and squid, with GenBank reference sequences showing high percentage matches in BLAST and congeners clustering monophyletically in the NJ and DO trees. Depending on the studied species and on the potential prey, the 16S mtDNA gene seems to be a reliable marker for dietary analysis of marine predators.

The incompleteness of reference databases has been widely acknowledged as the main factor limiting accurate taxonomic assignments using DNA barcodes (Valdez-Moreno *et al.* 2012), but is also a limiting factor in morphological analyses. In the case of the families Trichiuridae and Synaphobranchidae only a few species and genera are represented in the

399 GenBank database and, therefore, only family level assignments were obtainable. The
400 expansion of the taxonomic and geographic scope of fish and cephalopod reference material
401 in GenBank, particularly for oceanic species, is needed to disentangle the identification of
402 closely related species.

403 Prey identified in the families Exocoetidae and Carangidae produced very similar
404 BLAST matches to different genera and species in the GenBank database. In the case of
405 Exocoetidae, query sequences produced matches with percentage of similarity > 98% even
406 between different genera. Although most of our sequences could be reliably assigned to genus
407 level based on the trees, many species were polyphyletic suggesting that evolutionary
408 relationships in these genera are unclear (especially *Cheilopogon*). These findings may be an
409 artefact inherent to the use of single markers, which represent only a small snapshot of the
410 evolutionary history of species. However, studies using mitochondrial and nuclear markers
411 (cytb and RAG2) have also reported polyphyly of the genus *Cheilopogon* (Lewallen *et al.*
412 2011). Therefore, species assignments based on divergence thresholds should be interpreted
413 with caution in these groups. Regardless of the success in species identification, DNA
414 barcoding also allowed identification of a high number of MOTUs, reflecting the diversity of
415 flying-fishes around the Selvagens islands.

416 The complementary use of DNA and conventional methods allowed us to identify fish
417 vertebrae of particular species (e.g., *Ranzania laevis*, *Katsuwonus pelamis*) and to use those in
418 subsequent identification and quantification, bridging some of the gaps in our morphological
419 reference collection. We should note that it is not always possible to collect viable tissue
420 samples from digested prey remains (frequent in Procellariiform diet samples) in order to
421 perform genetic analysis. Therefore, relying on a combined approach, we were able to
422 maximise the identification and quantification of different types of prey.

Diet of non-breeders

The diet of non-breeding/immature seabirds has been seldom studied, mainly due to the difficulty of obtaining a sufficient number of samples at the breeding colony (Barrett *et al.* 2007; Granadeiro *et al.* 2009). Most researchers have relied on the isotopic analysis of tissues, and suggested that immatures feed at a lower trophic level (Forero *et al.* 2002; Votier *et al.* 2011), but the lack of taxonomic resolution of this approach prevented a better understanding of those differences.

At Selvagem Grande, a high number of non-breeding individuals, mostly composed by immature individuals, attend the colony during the breeding period (Granadeiro *et al.* 2009). We found that, during the chick-rearing period (August/September), the diet of non-breeders was substantially different from that of breeders, with a higher incidence of cephalopods (FO = 63.2% versus 35.1%) in the former group. Furthermore, in June of the same year, the incidence of cephalopods (mostly neon-flying squid) in the diet of non-breeders was even higher (FO = 93.9%). These results strongly point towards an ontogenetic shift in the trophic niche, the causes of which need to be evaluated by further studies. Non-breeders are less constrained by the need to attend the nesting colony and as such we would have expected them to feed more on distant (coastal) prey. However, the opposite pattern was revealed by our data, as squid in our system is more often captured in offshore waters (unpublished data). Does this differentiation reflect a difference in foraging abilities of breeders and non-breeders? Or could non-breeders be forced, by the competitively superior breeders, out of the rich feeding areas of the coastal upwelling (Ramos *et al.* 2013)? Our results urge more research in this area. Given the potential susceptibility of pelagic seabirds, such as the Cory's shearwater, to mortality linked to fishing vessels (Belda & Sanchez 2001) and to changes in

447 the availability of their prey (Paiva *et al.* 2013), these results have clear implications. They
448 suggest that different segments of seabird populations are likely to respond differently to
449 ecosystem changes, or to the impacts of human activities, and those need to be taken into
450 account, for example, in demographic modelling (Oro *et al.* 2010).
451
452 **The influence of sex on diet**
453 Direct evidence of sex-related dietary differences in pelagic seabirds is scarce (e.g.,
454 Xavier & Croxall 2005, Castillo-Guerrero *et al.* 2011) and most studies that investigated this
455 issue were based on a small number of samples (e.g., Zavalaga *et al.* 2007; Xavier *et al.*
456 2011). Despite that, many studies (mostly based on stable isotopes or tracking) clearly
457 showed the existence of a sex-related spatial or isotopic segregation in several seabird
458 populations, often linked to sexual dimorphism (Phillips *et al.* 2011). We found clear dietary
459 differences between sexes in Cory's shearwaters during the chick-rearing period, with males
460 feeding more on sardines and less on chub mackerel than females. Despite the marked
461 morphologic differences between sexes (Navarro *et al.* 2009; Ramos *et al.* 2009), sex-related
462 differences in the diet or in foraging ranges of Cory's shearwaters were not found in previous
463 studies (Navarro *et al.* 2007; Navarro *et al.* 2009; Xavier *et al.* 2011).
464 Male Cory's shearwaters are heavier, with larger bills and longer wings than females
465 (Navarro *et al.* 2009; Ramos *et al.* 2009). It is possible that the higher wing load of males
466 could provide them with greater mobility (Ramos *et al.* 2009) and enable them to increase
467 their foraging range, in relation to females. Indeed, Cory's shearwaters from Selvagem
468 Grande are known to prey on sardines mostly during long-distance foraging trips along the
469 African coast (unpublished data). Weimerskirch *et al.* (2006) also described a greater foraging
470 range of females in relation to males in red footed boobies *Sula sula*, presumably due the

larger size of females. However, sexual divergence in provisioning or foraging specialisation could also explain diet differences (Phillips *et al.* 2004) and this issue requires further investigation.

Inter-annual and seasonal variations in diet

There were inter-annual differences in the occurrence of some prey species in the diet of the shearwaters, namely sardine and cephalopods, which were more frequent in 2010. Cory's shearwaters are generalist predators (Thibault *et al.* 1997) and it is likely that these temporal variations may reflect a change in the abundance or availability of their main prey. However, inter-annual differences were smaller than variations linked to season and to foraging domain, found in this and in other studies (Paiva *et al.* 2010; Neves *et al.* 2012). Our results also contrast with previous studies at the Azores, where much more marked inter-annual variations in the consumption of fish and cephalopods were detected (Granadeiro *et al.* 1998; Paiva *et al.* 2010; Xavier *et al.* 2011; Neves *et al.* 2012). This suggests that the marine environment in the vicinity of the Selvagens Islands presented limited inter-annual changes in summer, which may be a general feature of these pelagic subtropical waters.

The diet of Cory's shearwaters was substantially different between the pre-laying and chick-rearing periods. During the pre-laying period, shearwaters fed mainly on trumpet fish and horse/blue jack mackerel. These prey species were of low importance during the chick-rearing period, when shearwaters increased the consumption of chub mackerel, sardine and pilot-fish. This variation in diet could be related to increased selectivity in prey choice, since parents are expected to select larger or higher-quality prey for their chicks (Wilson *et al.* 2004). Moreover, foraging areas explored by Cory's shearwaters are known to vary through

494 the breeding season (e.g., Navarro *et al.* 2007), possibly contributing to these striking seasonal
495 changes in diet.
496
497 **Conclusions**
498 Our study highlights the importance of combining different techniques to accurately
499 describe the diet of a pelagic seabird. The use of DNA barcoding and morphological analysis
500 proved to be very efficient to study the diet of Cory´s shearwaters, by improving both the
501 taxonomical resolution and the quantification of prey species. This approach is likely to be
502 useful in future seabird dietary studies. We also show the occurrence of trophic segregation
503 between birds of different breeding status and sex, highlighting the need to further investigate
504 the dietary choices of different population segments. Understanding the sources of dietary
505 variation within a seabird population will be important for instituting appropriate conservation
506 or population management measures.
507
508 **Acknowledgements**
509
510 Parque Natural da Madeira and in particular Paulo Oliveira, Dília Menezes and Carolina Santos, provided permissions and
511 logistical support to carry out the work at Selvagem Grande. We are grateful to all who gave us valuable help in field work:
512 Maria Dias, Ricardo Rocha, Ana Almeida, Filipe Moniz, Rui Rebelo, Teresa Catry, Ana Sofia Dias, Mariana Marques,
513 Miguel Lecoq, Eduardo Santos and Thijs Valkenburg. We are also thankful to the support provided by the wardens of the
514 Nature Reserve during our stays. This study was financed by Fundação para a Ciência e Tecnologia (FCT-Portugal) through
515 project PTDC/MAR/71927/2006, project PEst-OE/MAR/UI0331/2011 and Programa Ciência 2007, and through doctoral
516 fellowships to H. Alonso (BD/47055/2008) and S. Waap (BD/73656/2010).
517

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Barnett A, Redd KS, Frusher SD, Stevens JD, Semmens JM (2010) Non-lethal method to obtain stomach samples from a large marine predator and the use of DNA analysis to improve dietary information. *Journal of Experimental Marine Biology and Ecology*, **393**, 188–192.

Barrett RT, Camphuysen KCJ, Anker-Nilssen T, Chardine JW, Furness RW, Garthe S, Hüppop O, Leopold MF, Montevecchi WA, Veit RR (2007) Diet studies of seabirds: a review and recommendations. *ICES Journal of Marine Science*, **64**, 1675–1691.

Belda EJ, Sanchez A (2001) Seabird mortality on longline fisheries in the western Mediterranean: factors affecting bycatch and proposed mitigating measures. *Biological Conservation*, **98**, 357–363.

Bowen WD, Iverson SJ (2013) Methods of estimating marine mammal diets: A review of validation experiments and sources of bias and uncertainty. *Marine Mammal Science*, **29**, 719–754.

Bowser AK, Diamond AW, Addison JA (2013) From Puffins to Plankton: A DNA-Based Analysis of a Seabird Food Chain in the Northern Gulf of Maine. *PloS ONE*, **8**, e83152.

Castillo-Guerrero JA, Mellink E (2011) Occasional inter-sex differences in diet and foraging behavior of the Blue-footed Booby: maximizing chick rearing in a variable environment? *Journal of Ornithology*, **152**, 269–277.

Dai L, Zheng X, Kong L, Li Q (2012) DNA barcoding analysis of Coleoidea (Mollusca: Cephalopoda) from Chinese waters. *Molecular Ecology Resources*, **12**, 437–447.

Daunt F, Wanless S, Harris MP, Money L, Monaghan P (2007) Older and wiser: improvements in breeding success are linked to better foraging performance in European shags. *Functional Ecology*, **21**, 561–567.

Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*, **11**, 2039–2048.

Deagle BE, Gales NJ, Evans K, Jarman SN, Robinson S, Trebilco R, Hindell MA (2007) Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLoS ONE*, **2**, e831.

De Laet (2010) Letter to the editor. A problem in POY tree searches (and its work-around) when some sequences are observed to be absent in some terminals. *Cladistics*, **26**, 453–455.

den Hartog JC , Clarke MR (1996) A study of stomach contents of Cory's Shearwater, *Calonectris diomedea borealis* (Cory, 1881) (Aves: Procellariidae), from the Macaronesian Islands. *Zoologische Mededeelingen*, **70**, 117–133.

Dunn MR, Szabo A, McVeagh MS, Smith PJ (2010) The diet of deepwater sharks and the benefits of using DNA identification of prey. *Deep Sea Research Part I*, **57**, 923–930.

Forero MG, Hobson KA, Bortolotti GR, Donázar JA, Bertellotti M, Blanco G (2002) Food resource utilisation by the Magellanic penguin evaluated through stable-isotope analysis: segregation by sex and age and influence on offspring quality. *Marine Ecology Progress Series*, **234**, 289–299.

Goss-Custard JD, Durell SLVD, McGrorty S, Reading CJ (1982) Use of mussel *Mytilus edulis* beds by oystercatchers *Haematopus ostralegus* according to age and population size. *Journal of Animal Ecology*, **51**, 543–554.

Granadeiro JP (1993) Variation in measurements of Cory's Shearwater between populations and sexing by discriminant analysis. *Ringing and Migration*, **14**, 103–112.

Granadeiro JP, Alonso H, Almada V, Menezes D, Phillips RA, Catry P (2009) Mysterious attendance cycles in Cory's shearwater *Calonectris diomedea*: an exploration of patterns and hypotheses. *Animal Behaviour*, **78**, 1455–1462.

Granadeiro JP, Monteiro LR, Furness RW (1998) Diet and feeding ecology of Cory's shearwater *Calonectris diomedea* in the Azores, north-east Atlantic. *Marine Ecology Progress Series*, **166**, 267–276.

Granadeiro JP, Dias MP, Rebelo R, Santos CD, Catry P (2006) Numbers and population trends of Cory's Shearwater *Calonectris diomedea* at Selvagem Grande, northeast Atlantic. *Waterbirds*, **29**, 56–60.

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

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London - Biological Sciences,* **270**, 313–321.

Jarman SN, Gales NJ, Tierney M, Gill PC, Elliott NG (2002) A DNA-based method for identification of krill species and its application to analysing the diet of marine vertebrate predators. *Molecular Ecology*, **11**, 2679–2690.

Jarman SN, McInnes JC, Faux C, Polanowski AM, Marthick J, Deagle BE, Southwell C, Emmerson L (2013) Adélie Penguin Population Diet Monitoring by Analysis of Food DNA in Scats. *PloS ONE*, **8**, e82227.

Jedlicka JA, Sharma AM, Almeida RP (2013) Molecular tools reveal diets of insectivorous birds from predator fecal matter. *Conservation Genetics Resources*, **5**, 879–885.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kitowski I (2003) Age-related differences in foraging behavior of Montagu's harrier *Circus pygargus* males in south-east Poland. *Acta Ethologica*, **6**, 35–38.

Kochzius M, Seidel C, Antoniou A, Botla SK, Campo D, Cariani A, Vazquez EG, Hauschild J, Hervet C, Hjorleifsdóttir S, Hreggvidsson G, Kappel K, Landi M, Magoulas A, Marteinsson V, Nolte M, Planes S, Tinti F, Turan C, Venugopal MN, Weber H, Blohm D (2010) Identifying fishes through DNA barcodes and microarrays. *PLoS ONE,* **5**, e12620.

Lewallen EA, Pitman RL, Kjartanson SL, Lovejoy NR (2011) Molecular systematics of flyingfishes (Teleostei: Exocoetidae): evolution in the epipelagic zone. *Biological Journal of the Linnean Society,* **102**, 161–174.

Lucifora LO, García VB, Menni RC, Escalante AH, Hozbor NM (2009) Effects of body size, age and maturity stage on diet in a large shark: ecological and applied implications. *Ecological Research*, **24**, 109–118.

Moritz C, Cicero C (2004) DNA Barcoding: Promise and Pitfalls. *PLoS Biology*, **2**, e354.

Mougin J-L, Jouanin C, Roux F (1997) Intermittent breeding in Cory's Shearwater *Calonectris diomedea* of Selvagem Grande, North Atlantic. *Ibis*, **139**, 40–44.

Naito Y, Costa DP, Adachi T, Robinson PW, Fowler M, Takahashi A (2013) Unravelling the mysteries of a mesopelagic diet: a large apex predator specializes on small prey. *Functional Ecology*, **27**, 710–717.

Navarro J, González-Solís J, Viscor G (2007) Nutritional and feeding ecology in Cory's shearwater *Calonectris diomedea* during breeding. *Marine Ecology Progress Series*, **351**, 261–271.

Navarro J, Kaliontzopoulou A, Gonzalez-Solis J (2009) Sexual dimorphism in bill morphology and feeding ecology in Cory's shearwater (*Calonectris diomedea*). *Zoology*, **112**, 128–138.

Neves VC, Nolf D, Clarke M (2012) Spatio-temporal variation in the diet of Cory's shearwater *Calonectris diomedea* in the Azores archipelago, north-east Atlantic. *Deep Sea Research Part I*, **70**, 1–13.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2011) vegan: Community Ecology Package. R Package Version 2.0-2. http://CRAN.R-project.org/package=vegan

Oro D, Torres R, Rodríguez C, Drummond H (2010) Climatic influence on demographic parameters of a tropical seabird varies with age and sex. *Ecology*, **91**, 1205–1214.

Paiva VH, Xavier JC , Geraldes P, Ramírez I, Garthe S, Ramos JA (2010) Foraging ecology of Cory's shearwaters in different oceanic environments of the North Atlantic. *Marine Ecology Progress Series*, **410**, 257–268.

Paiva VH, Geraldes P, Marques V, Rodríguez R, Garthe S, Ramos JA (2013) Effects of environmental variability on different trophic levels of the North Atlantic food web. *Marine Ecology Progress Series*, **477**, 15–28.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: Hillis DM, Mable BK, Moritz C (eds) Molecular Systematics. Sinauer, Sunderland, MA, pp 205–247.

Phillips RA, McGill RAR, Dawson DA, Bearhop S (2011) Sexual segregation in distribution, diet and trophic level of seabirds: insights from stable isotope analysis. *Marine Biology*, **158**, 2199–2208.

Phillips RA, Silk JRD, Phalan B, Catry P, Croxall JP (2004) Seasonal sexual segregation in two *Thalassarche* albatross species: competitive exclusion, reproductive role specialization or foraging niche divergence? *Proceedings of the Royal Society of London - Biological Sciences*, **271**, 1283–1291.

Polis GA (1984) Age structure component of niche width and intraspecific resource partitioning: can age groups function as ecological species? *American Naturalist*, **123**, 541–564.

Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SD, Taberlet P (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, **21**, 1931–1950.

Quéro J-C, Porché P, Vayne JJ (2003) Guide des Poissons de l'Atlantique Européen. Delachaux & Niestlé, Paris.

R Development Core Team (2010) R: A Language and Environment for Statistical Computing (2.10.1). R Foundation for Statistical Computing, Vienna

Ramos JA, Granadeiro JP, Phillips RA, Catry P (2009) Flight morphology and foraging behaviour of male and female Cory's Shearwaters. *Condor*, **111**, 424–432.

Ramos R, Granadeiro JP, Rodríguez B, Navarro J, Paiva VH, Bécares J, Reyes-González JM, Fagundes I, Ruiz A, Arcos P, González-Solís J, Catry P (2013) Meta-population feeding grounds of Cory's shearwater in the subtropical Atlantic Ocean: implications for the definition of Marine Protected Areas based on tracking studies. *Diversity and Distributions*, **19**, 1284–1298.

Symondson WOC (2002) Molecular identification of prey in predator diets. *Molecular Ecology*, **11**, 627–641.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA 6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Thibault J-C, Bretagnolle V, Rabouam C (1997) *Calonectris diomedea* Cory's Shearwater. *Birds of the Western Palearctic Update*, **1**, 75–98.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Tuset VM, Lombarte A, Assis CA (2008) Otolith atlas for the western Mediterranean, north and central eastern Atlantic. *Scientia Marina*, **72**, 7–198.

Valdez-Moreno M, Quintal-Lizama C, Gómez-Lozano R, García-Rivas MdC (2012) Monitoring an alien invasion: DNA barcoding and the identification of lionfish and their prey on coral reefs of the Mexican caribbean. *PLoS ONE*, **7**, e36636.

Varón A, Vinh LS, Wheeler WC (2010) POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*, **26**, 72–85.

Votier SC, Grecian WJ, Patrick S, Newton J (2011) Inter-colony movements, at sea behaviour and foraging in an immature seabird: results from GPS-PTT tracking, radiotracking and stable isotope analysis. *Marine Biology*, **158**, 355–362.

Weimerskirch H, Le Corre M, Ropert-Coudert Y, Kato A, Marsac F (2006) Sex-specific foraging behaviour in a seabird with reversal sexual dimorphism: the red-footed booby. *Oecologia*, **146**, 681–691.

Wilson RP (1984) An improved stomach pump for penguins and other seabirds. *Journal of Field Ornithology*, **55**, 109–112.

Wilson LJ, Daunt F, Wanless S (2004) Self-feeding and chick-provisioning diet differ in the common guillemot *Uria aalge*. *Ardea*, **92**, 197–208.

Wilson JJ, Rougerie R, Schonfeld J, Janzen DH, Hallwachs W, Hajibabaei M, Kitching IJ, Haxaire J, Hebert PDN (2011) When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. *BMC Ecology,* **11**, 1–14.

Xavier JC, Croxall JP (2005) Sexual differences in foraging behaviour and diets: a case study of wandering albatrosses. In: Ruckstuhl KE, Neuhaus P (eds) Sexual segregation in vertebrates: ecology of the two sexes. Cambridge University Press, Cambridge, pp 74–91.

Xavier JC, Magalhães MC, Mendonça AS, Antunes M, Carvalho N, Machete M, Santos RS, Paiva V, Hamer KC (2011) Changes in diet of Cory's Shearwaters *Calonectris diomedea* breeding in the Azores. *Marine Ornithology*, **39**, 129–134.

Zavalaga CB, Benvenuti S, Dall'Antonia L, Emslie SD (2007) Diving behavior of blue-footed boobies *Sula nebouxii* in northern Peru in relation to sex, body size and prey type. *Marine Ecology Progress Series*, **336**, 291–303.

Zhang J, Hanner R (2012) Molecular approach to the identification of fish in the South China Sea. *PLoS ONE*, **7**, e30621.

Data accessibility

Sequence data have been deposited to GenBank (accession numbers KC603479–KC603537). Input files for POY and Mega 6 have been deposited in Dryad.

Authors contributions

JPG, PC, HA and SW designed the study; PC, JPG, JAR and WOCS provided guidance during the study; HA, JPG and PC collected the data in the field; HA and JX identified and quantified prey trough conventional methods; HA and JPG analysed the data; SW performed the molecular part of the study; HA, SW, PC, JPG, and WOCS wrote the manuscript.

Supplementary information

Additional supporting information may be found in the online version of this article.

Figure S1 NJ tree based method for assignment of morphologically unidentified specimens in Cory's shearwater diet.

Figure S2 NJ tree based method for assignment of morphologically unidentified specimens in Cory's shearwater diet. .

Table S1 Genbank accession numbers of prey identified using DNA barcoding.

689 **Figures and tables**

690
691 **Fig. 1** Tree estimated in POY for identification of teleosts using direct optimisation (DO)
692 method. Query sequences and Genbank accession numbers of morphologically unidentified
693 specimens for which no reliable identification could be obtained in BLAST are shown
694 (unidentified specimens code). Representative genera of the families that produced the nearest
695 match are included in the tree. Query sequences clustering with *Cheilopogon pinnatibarbatus*,
696 *C. melanurus* and *Trachurus picturatus* corroborate previous BLAST results (100% of
697 similarity). The tree shows the number of different MOTUs (different prey) obtained in each
698 family. Nodal support is presented for bootstrap values ≥ 70.
699
700 **Fig. 2** Diet of Cory's shearwaters *Calonectris diomedea* in the pre-laying (30 diet samples
701 with 318 prey) and chick-rearing (188 diet samples with 553 prey) periods of 2010: a)
702 Frequency of occurrence of each prey type (%), b) Numerical frequency of each prey type
703 (%).
704
705 **Fig. 3** Diets of breeders (N = 248 diet samples) and non-breeders (N = 19) among Cory's
706 shearwaters *Calonectris diomedea* during the chick-rearing period of 2009 (Frequency of
707 occurrence, %).
708
709

710 **Table 1** Cory´s shearwater (*Calonectris diomedea*) prey identified using DNA barcoding of
711 16S mtDNA.

	Family	Genus/Species	Specimens	Percentage of Similarity	Phylogenetic analysis
Teleostei	Carangidae	*Trachurus* sp.	5		**
	Carangidae	*Trachurus picturatus*	2	100	***
	Coryphaenidae	*Coryphaena equiselis**	1	100	
	Diretmidae	*Diretmus argenteus**	1	99.8[a]	
	Exocoetidae	*Cheilopogon melanurus**	1	100	***
	Exocoetidae	*Cheilopogon pinnatibarbatus**	2	100	***
	Exocoetidae	*Cheilopogon* sp.	1		**
	Exocoetidae	*Exocoetus* sp.	16		**
	Exocoetidae	Unidentified	2		
	Halosauridae	*Halosaurus* sp.*	1		**
	Molidae	*Ranzania laevis**	2	100	
	Myctophidae	*Diaphus* sp.	1		**
	Myctophidae	*Lampadena atlantica**	1	100	
	Neoscopelidae	*Neoscopelus macrolepidotus**	1	100	
	Scombridae	*Katsuwonus pelamis**	2	100	
	Sparidae	*Boops boops**	2	99.1[b]	
	Sternoptychidae	*Argyropelecus* sp.*	1		**
	Synaphobranchidae*	Unidentified	2		**
	Trichiuridae*	Unidentified	2		**
Cephalopods	Chiroteuthidae	*Chiroteuthis mega**	1	99.8[b]	
	Cranchiidae	*Taonis pavo*	1	100	
	Histioteuthidae	*Histioteuthis* sp.	1		**
	Ommastrephidae	*Ommastrephes*	8		**

712 Single (*) asterisk correspond to taxa not recorded previously in the diet of Cory´s shearwater (den Hartog & Clarke 1996,
713 Granadeiro *et al.* 1998, Paiva *et al.* 2010, Xavier *et al.* 2011, Neves *et al.* 2012). Double (**) and triple (***) asterisks
714 represent positive genus and species assignments based on the Neighbour joining (NJ) and DO trees. Similarity percentages
715 with the GenBank reference sequences for species identifications using BLAST are shown (a) Monotypic species, (b)
716 assignment based on the geographical distribution of the taxa

717

718 **Table 2** Frequency of occurrence (FO %) and numerical frequency (NF %) of prey, identified
719 by a combined use of morphologic analysis and DNA barcoding, in the diet of Cory's
720 shearwaters (*Calonectris diomedea*). Diet samples were collected in two different periods,
721 pre-laying (only in 2010) and chick-rearing (in 2008, 2009 and 2010), at Selvagem Grande.
722 Number of samples and prey is presented in brackets.

	Pre-laying		Chick-rearing					
	2010		2008		2009		2010	
	FO (30)	NF (318)	FO (180)	NF (416)	FO (248)	NF (631)	FO (188)	NF (553)
CEPHALOPODA	13.3	1.3	27.2	13.9	35.1	24.9	46.8	22.8
Chiroteuthidae								
*Chiroteuthis mega**	3.3	0.3						
Chiroteuthis sp.					1.2	0.5		
Cranchiidae								
Taonius pavo	3.3	0.3	0.6	0.2			1.1	0.5
Cranchia sp.								
Grimalditeuthidae								
Grimalditeuthis bonplandi			0.6	0.2				
Histioteuthidae								
Histioteuthis arcturi			1.1	0.5	1.2	0.5	5.3	2.0
Histioteuthis meleagroteuthis							0.5	0.2
Histioteuthis sp.			1.1	0.5	0.8	0.3		
Mastigoteuthidae								
Mastigoteuthis sp.					0.4	0.2		
Unidentified							2.1	0.7
Neoteuthidae								
Neoteuthis sp.							0.5	0.2
Ommastrephidae								
*Ommastrephes bartramii**	3.3	0.3	7.7	3.8	13.7	9.2	6.4	2.7
Octopoteuthidae								
Taningia danae			0.6	0.2	0.4	0.2		
Onychoteuthidae								
Ancistroteuthis lichtensteinii					0.4	0.2		
Sepiidae								
Unidentified							0.5	0.2

Unidentified cephalopods	3.3	0.3	17.2	8.4	25.8	13.9	36.2	16.3
FISH	86.7	98.7	91.1	86.1	88.7	75.1	89.9	76.9
Belonidae								
Belone belone			2.2	1.0	0.4	0.2	3.7	1.6
Caproidae								
Capros aper					0.4	0.2	0.5	0.2
Carangidae								
Naucrates ductor			15.6	15.9	16.1	14.9	13.3	10.7
*Trachurus picturatus**			0.6	0.2	0.4	0.2		
Trachurus sp.	40.0	9.7	1.7	0.7	2.8	1.4	4.8	4.0
Clupeidae								
Sardina pilchardus			7.8	4.1	9.3	4.3	20.2	13.2
Sardinella sp.					0.4	0.2		
Congridae								
Conger conger					0.8	0.3	0.5	0.2
Coryphaenidae								
*Coryphaena equiselis**					0.4	0.2		
Coryphaena sp.			0.6	0.2			1.1	0.5
Diretmidae								
*Diretmus argenteus**							2.7	0.9
Engraulidae								
Engraulis encrasicolus			1.7	1.0	3.6	4.0	1.1	0.5
Exocoetidae								
Exocoetus volitans			1.7	0.7				
Exocoetus sp.			8.3	4.1	1.2	0.5	4.8	1.6
Cheilopogon exsiliens					1.2	0.5		
*Cheilopogon melanurus**					0.4	0.2		
*Cheilopogon pinnatibarbatus**					0.4	0.2		
Cheilopogon sp.			1.1	0.5	1.2	0.5	1.1	0.4
Unidentified	3.3	0.3	6.7	2.9	4.0	2.5	2.7	0.9
Halosauridae								
Halosaurus sp.*			0.6	0.2	0.4	0.2		
Unidentified			0.6	0.5			0.5	0.2
Macroramphosidae								

Macroramphosus scolopax	36.7	83.3					1.1	0.4
Molidae								
*Ranzania laevis**							3.7	1.3
Myctophidae								
Diaphus sp.*					0.4	0.2		
*Lampadena atlantica**							0.5	0.2
Unidentified			1.7	1.0			1.6	0.9
Neoscopelidae								
*Neoscopelus macrolepidotus**			0.6	0.2				
Scomberesocidae								
Scomberesox sp.	3.3	0.3	2.2	1.7	1.2	0.5	4.3	2.9
Scombridae								
*Scomber colias/*sp.	23.3	3.8	46.7	39.2	51.2	36.7	35.6	23.3
*Katsuwonus pelamis**			0.6	0.2	3.2	1.3	2.7	0.9
Sparidae								
*Boops boops**					0.4	0.2	1.1	0.5
Sternoptychidae								
Argyropelecus sp.*			0.6	0.5				
Synaphobranchidae*								
Unidentified			1.1	0.5	1.2	0.5	3.7	1.4
Trichiuridae*								
Unidentified			4.4	1.9	1.2	0.5	4.8	1.8
Unidentified fish	13.3	0.9	19.4	8.9	13.3	5.5	22.3	8.3
CRUSTACEA			4.4		6.5		14.4	
Decapoda	3.3	0.3					0.5	0.4
Isopoda			1.1		3.6		11.1	
Unidentified crustacean	13.3		3.3		2.8		2.7	
INSECTA					0.8			
Halobatidae					0.8			

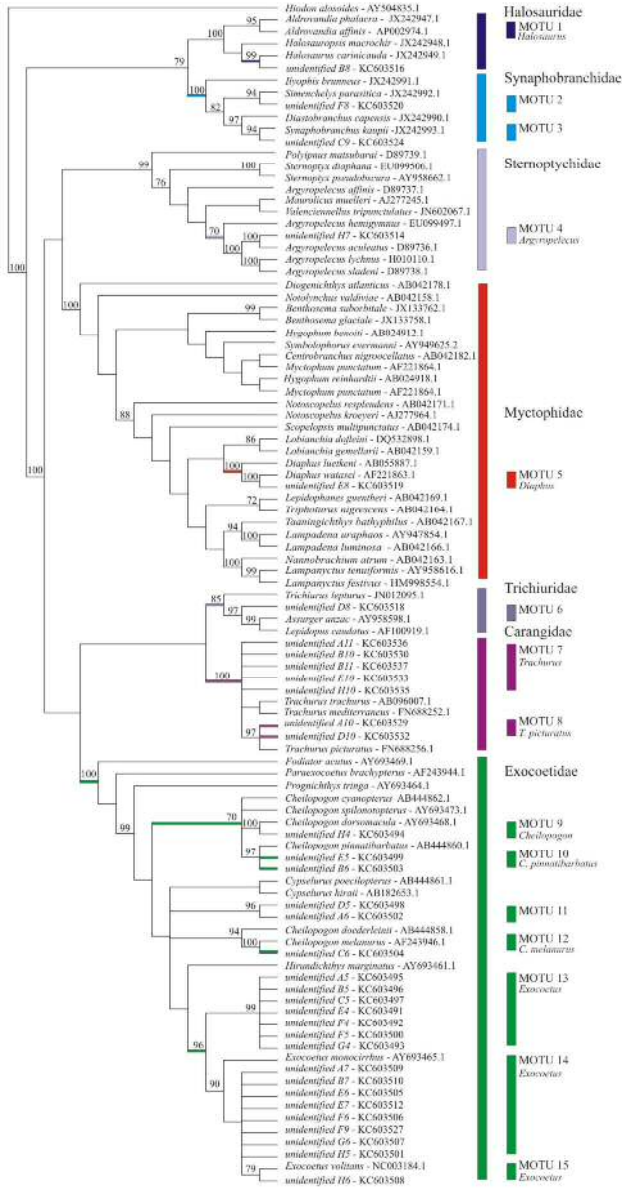
723 Single (*) asterisk represent taxa first identified through DNA barcoding of the 16S mtDNA.

724
725

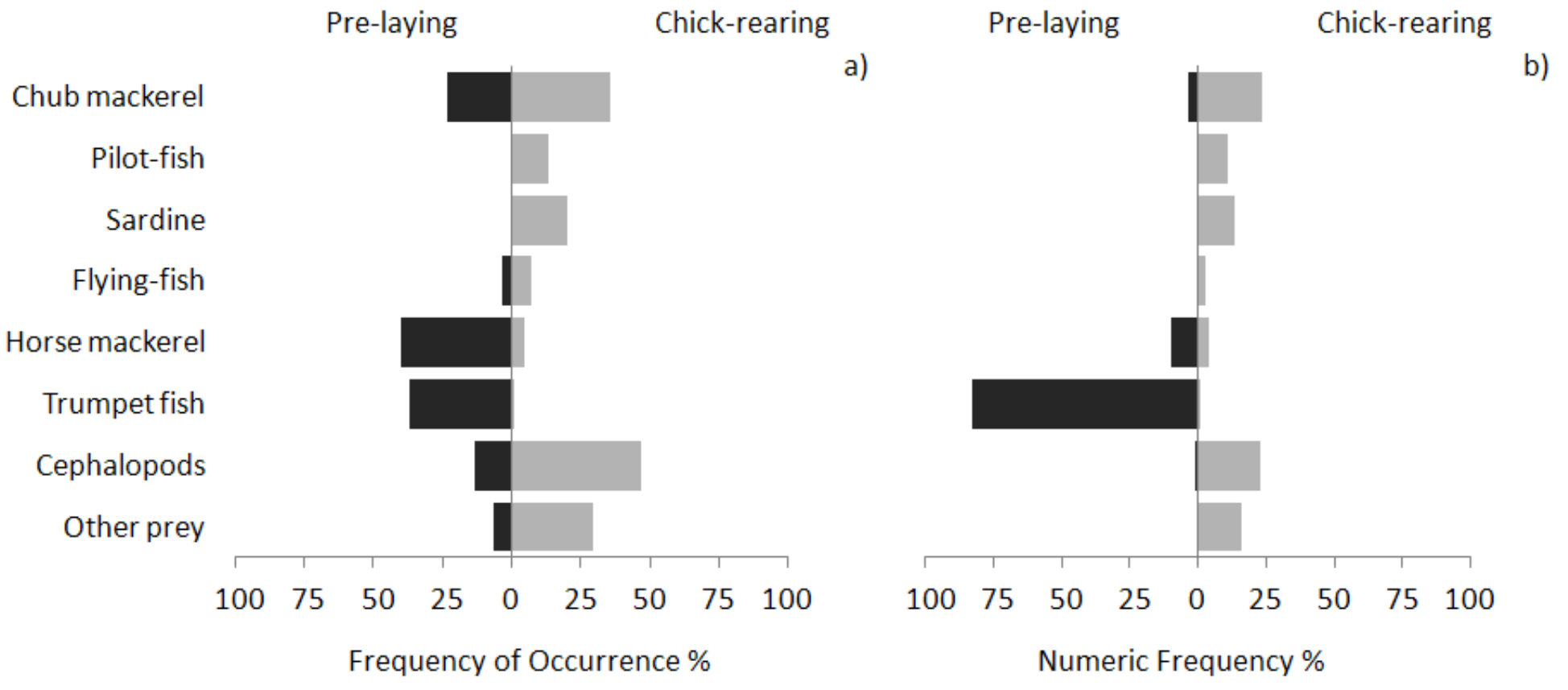
726 **Table 3** Frequency of occurrence (FO %) of prey in the diet of male and female Cory's
727 shearwaters (*Calonectris diomedea*), during the chick-rearing period of 2008, 2009 and 2010,
728 at Selvagem Grande island. Sample size is presented in brackets. Differences among sexes
729 were tested using a binomial GLM with a logit link function, controlling for the effect of year.

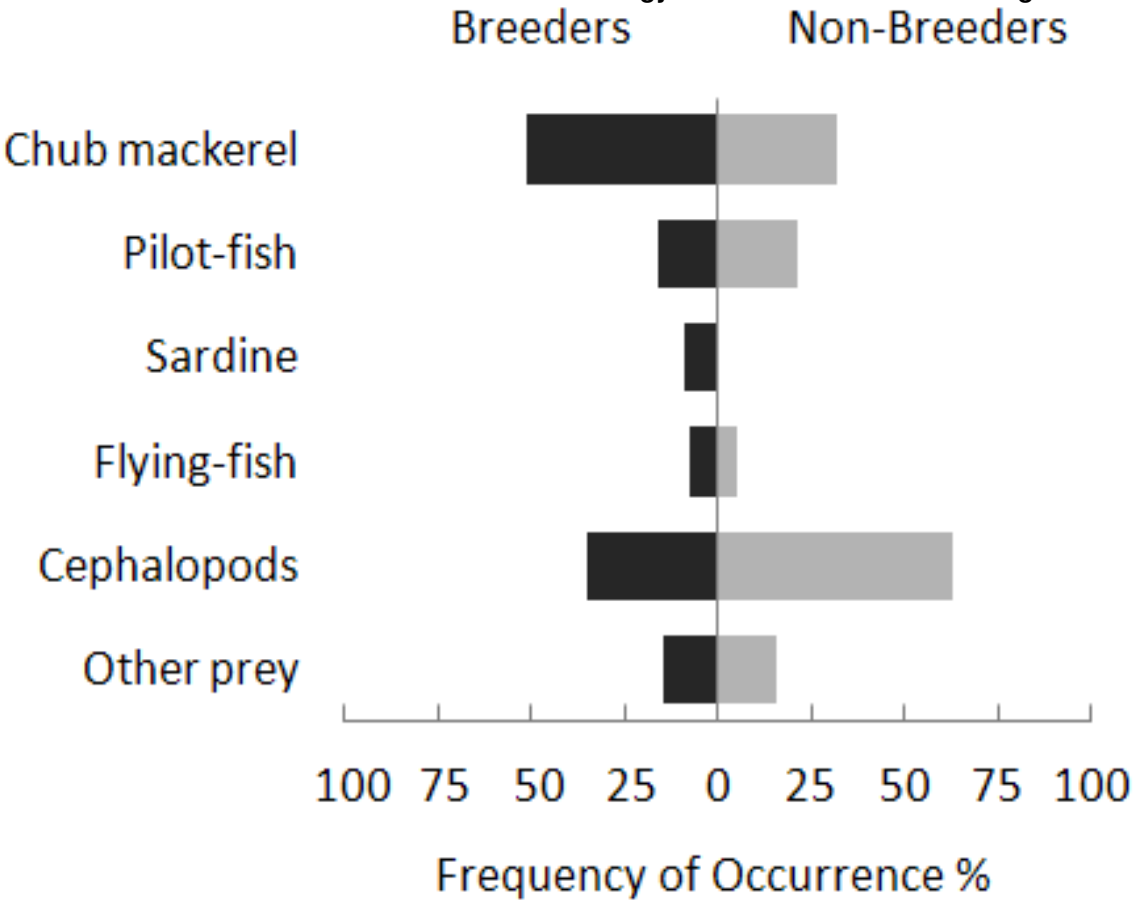
	Males (316)	Females (191)	Sex		Year	
	FO (%)	FO (%)	F	p	F	p
Fish	88.0 (278)	94.2 (180)	5.3	**< 0.05**	0.08	0.92
Scomber colias/sp.	38.0 (120)	58.6 (112)	21.7	**< 0.001**	5.9	**< 0.01**
Naucrates ductor	15.8 (50)	13.6 (26)	0.5	0.50	0.6	0.55
Sardina pilchardus	16.1 (51)	8.9 (17)	5.5	**< 0.05**	6.0	**< 0.01**
Trachurus sp.	4.7 (15)	2.1 (4)	2.3	0.13	0.4	0.69
Exocoetidae	10.1 (32)	5.8 (11)	2.9	0.09	0.9	0.41
Cephalopods	35.4 (112)	34.0 (65)	0.10	0.74	8.9	**< 0.001**

730



Tree estimated in POY for identification of teleosts using direct optimisation (DO) method.


Pre-laying
Chick-rearing
a)
Chub mackerel
Pilot-fish
Sardine
Flying-fish
Horse mackerel
Trumpet fish
Cephalopods
Other prey
100 75 50 25 0 25 50 75 100
Frequency of Occurrence %
Pre-laying
Chick-rearing
b)
100 75 50 25 0 25 50 75 100
Numeric Frequency %


Breeders
Non-Breeders
Chub mackerel
Pilot-fish
Sardine
Flying-fish
Cephalopods
Other prey
100 75 50 25 0 25 50 75 100
Frequency of Occurrence %

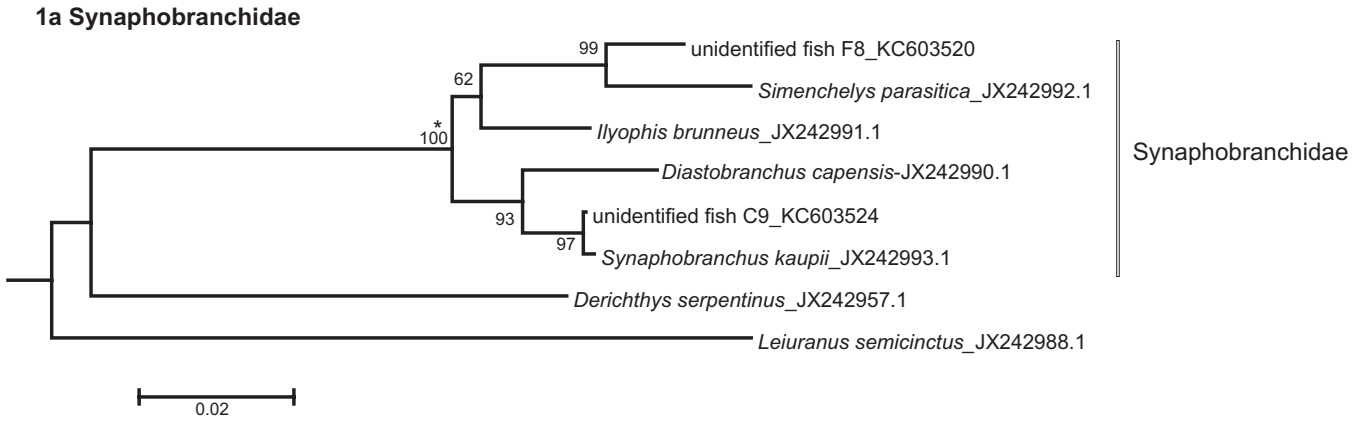


Figure S1. NJ tree based method for assignment of morphologically unidentified specimens in Cory's shearwater diet. Positive family level identifications were obtained for the families (a) Synaphobranchidae and (b) Trichiuridae. Specimens were positively assigned when monophyletic clusters with con-genera of the family were obtained. (*) Represent positive assignments. Bootstrap values are above nodes and are presented for a cut-off value > 50.

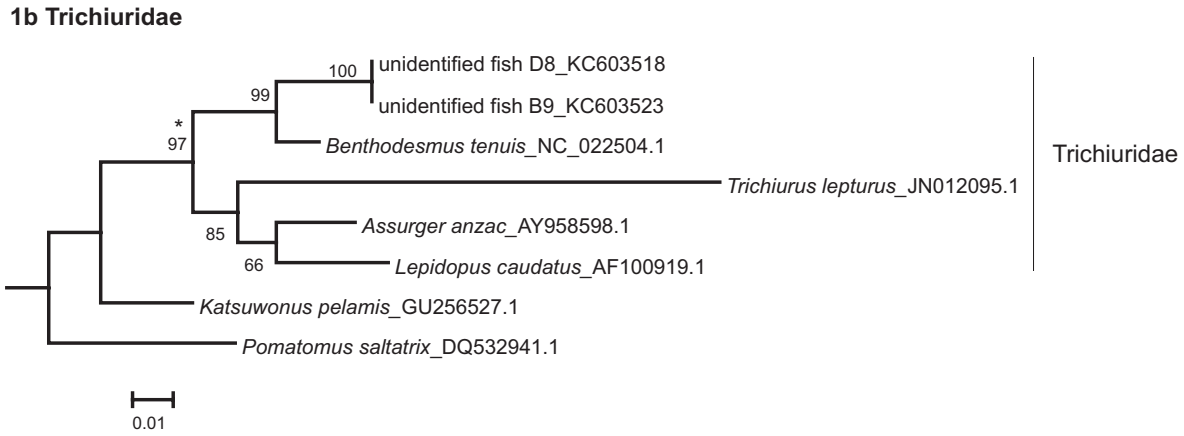


Figure S1. Continued

2a Carangidae: *Trachurus, T. picturatus*

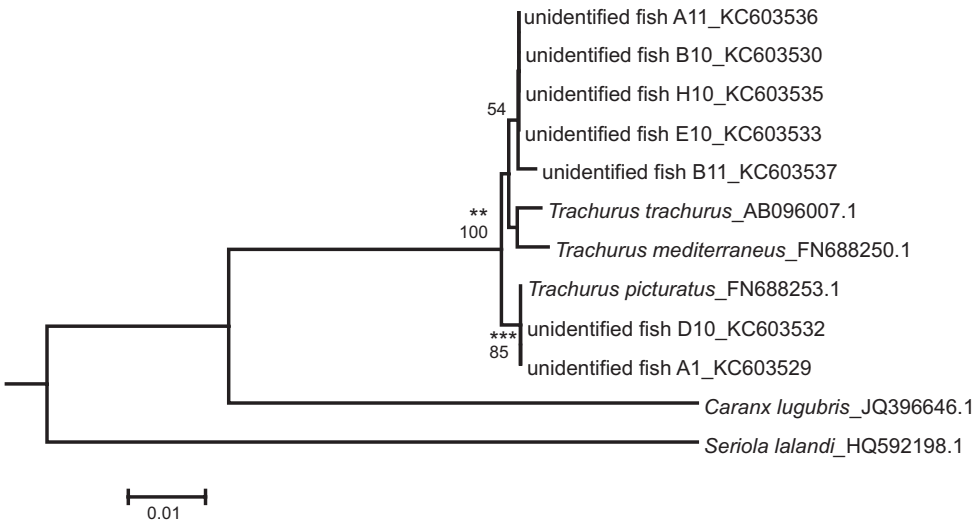


Figure S2. NJ tree based method for assignment of morphologically unidentified specimens in Cory's shearwater diet. Positive genus and species level identifications were obtained within the teleost families a) Carangidae b) Exocoetidae c) Halosauridae (d) Myctophidae (e) Sternoptychidae and cephalopods: f) Histioteuthidae g) Ommastrephidae. Specimens were positively assigned according to a strict criterion. (**) and (***) represent positive genus and species level assignments, respectively. Bootstrap values are above nodes and are presented for a cut-off value > 50

2b Exocoetidae: *Exocoetus sp, Cheilopogon sp, C. pinnatibarbatus, C. melanurus*

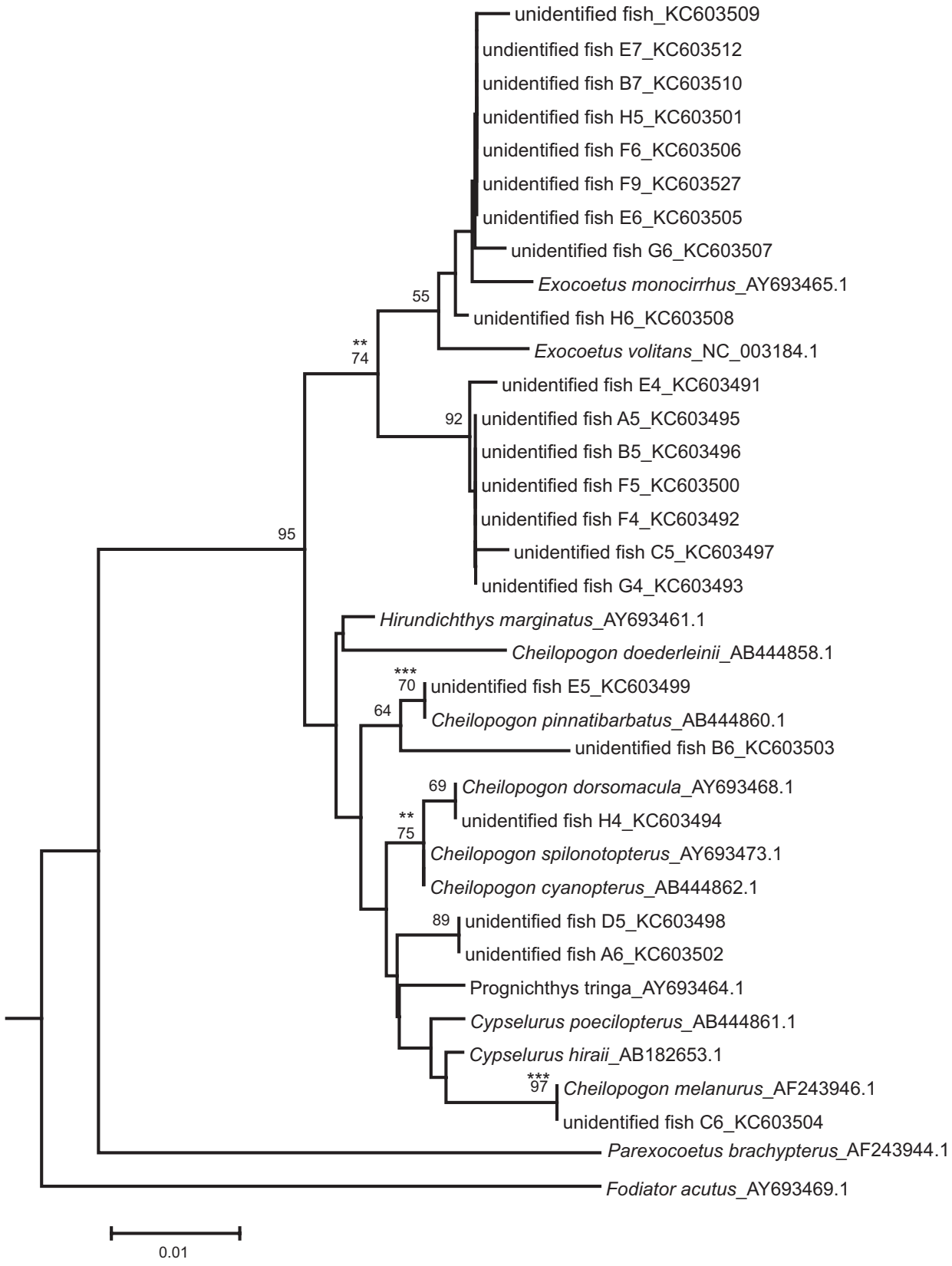


Figure S2. Continued

2c Halosauridae: *Halosaurus sp*

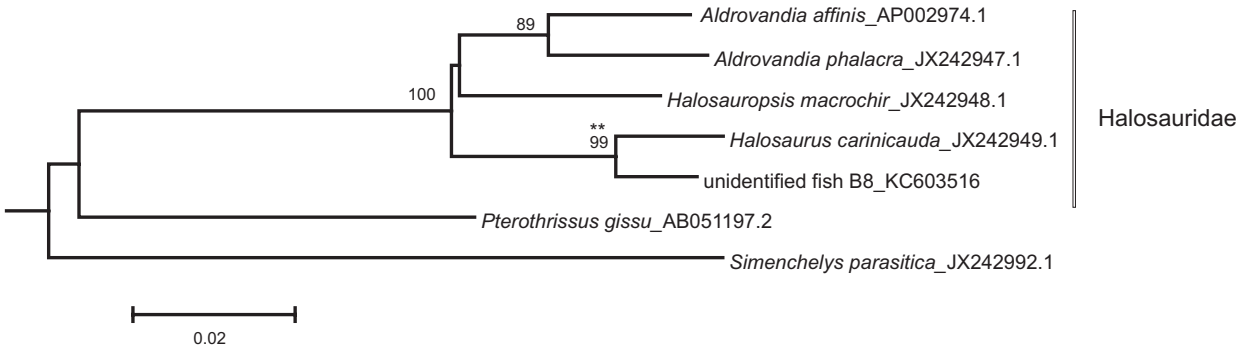


Figure S2. Continued

2d Myctophidae: *Diaphus sp*

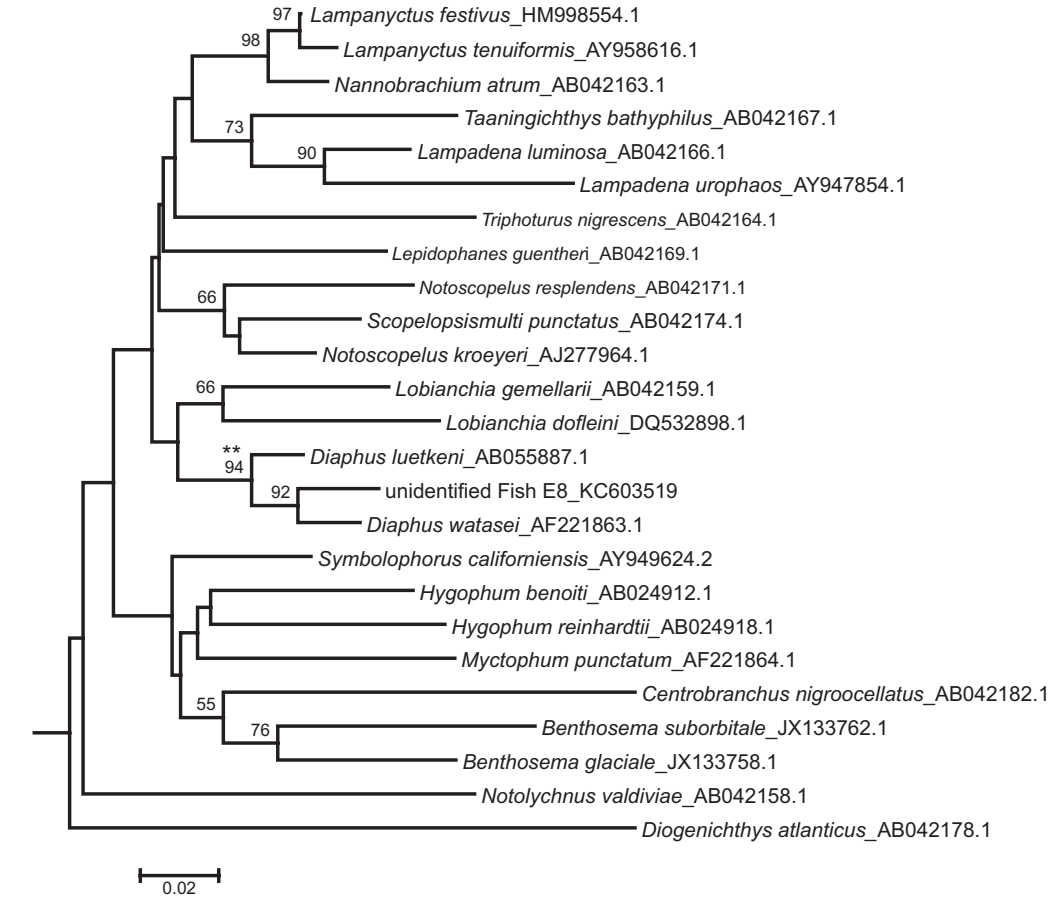


Figure S2. Continued

2e Sternoptychidae: *Argyropelecus sp*

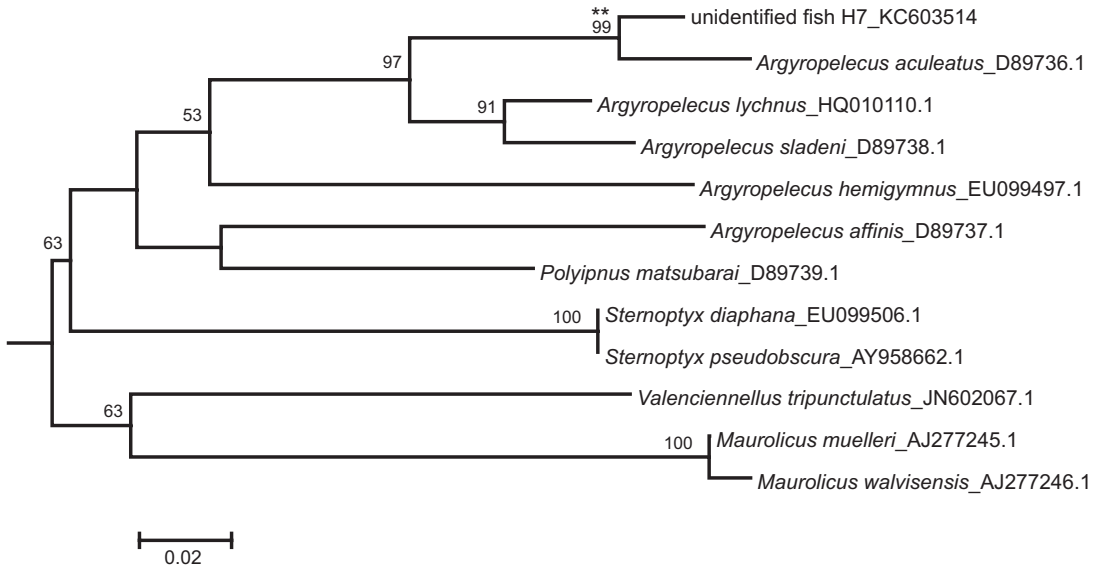


Figure S2. Continued

2f Histioteuthidae: *Histioteuthis* sp

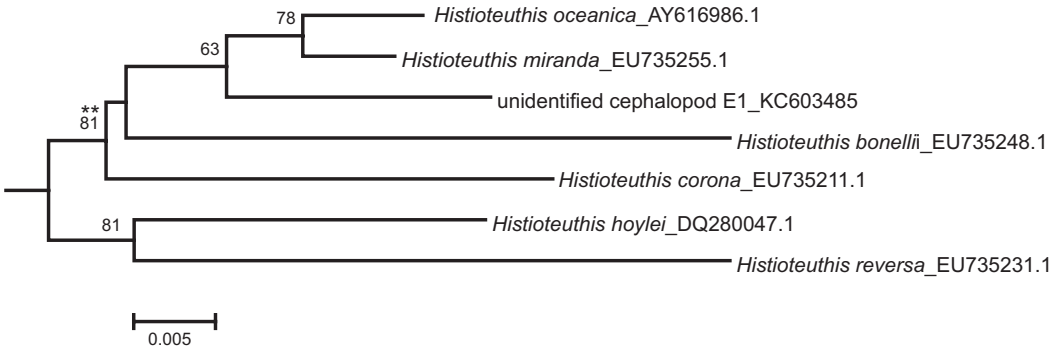


Figure S2. Continued

2g Ommastrephidae: *Ommastrephes bartramii*

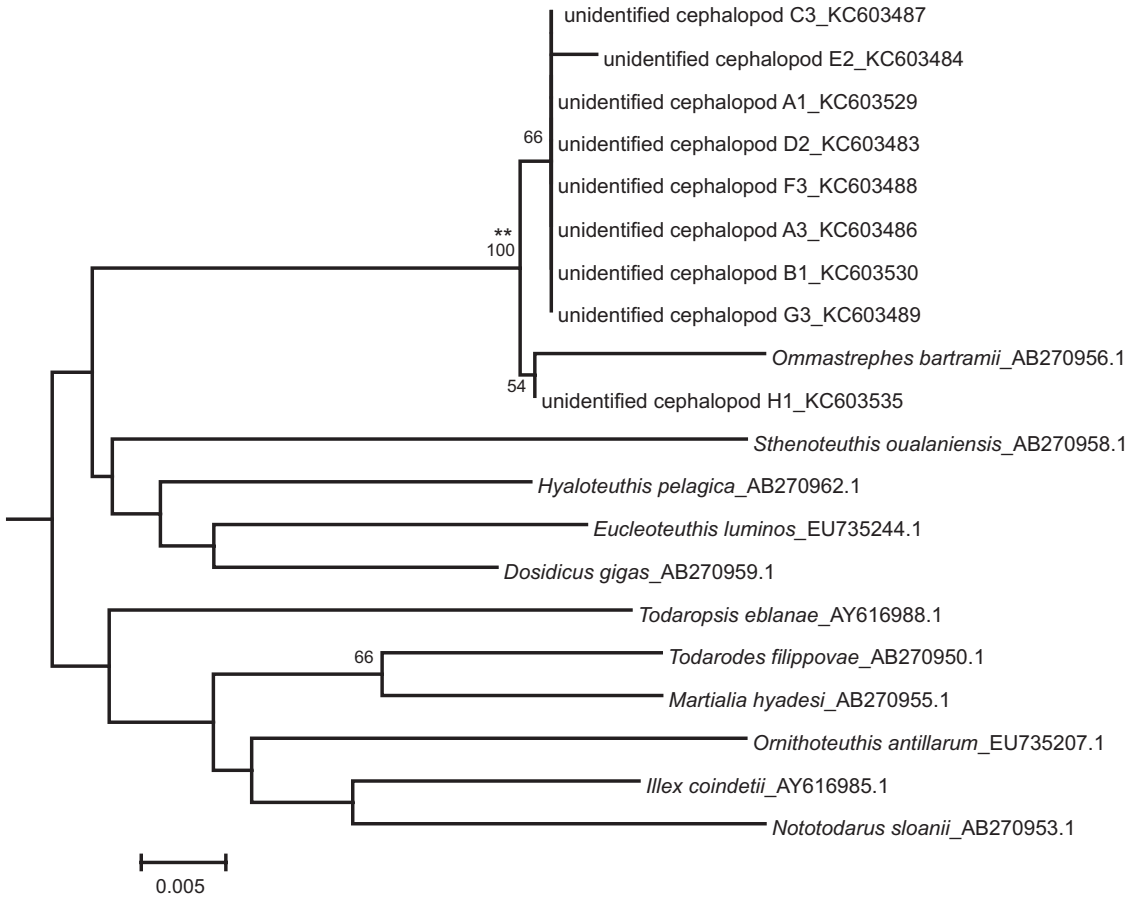


Figure S2. Continued

Table S1 Genbank accession numbers of prey identified using DNA barcoding. NI corresponds to unidentified prey taxon

Family	Genus	Species	Number of individuals	Genbank accession number
Histioteuthidae	*Histioteuthis*	NI	1	KC603485
Ommastrephidae	*Ommastrephes*	*Ommastrephes bartrammi*	9	KC603479, KC603480, KC603482-KC603484, KC603486-KC603489
Chiroteuthidae	*Chiroteuthis*	*Chiroteuthis mega*	1	KC603490
Cranchiidae	*Taonius*	*Taonius pavo*	1	KC603481
Carangidae	*Trachurus*	*Trachurus picturatus*	2	KC603529, KC603532
Carangidae	*Trachurus*	NI	5	KC603530, KC603533, KC603535-KC603537
Coryphaenidae	*Coryphaena*	*Coryphaena equiselis*	1	KC603517
Diretmidae	*Diretmus*	*Diretmus argenteus*	1	KC603521
Exocoetidae	*Cheilopogon*	NI	1	KC603494
Exocoetidae	*Cheilopogon*	*Cheilopogon melanurus*	1	KC603504
Exocoetidae	*Cheilopogon*	*Cheilopogon pinnatibarbatus*	1	KC603499
Exocoetidae	*Exocoetus*	NI	16	KC603491-KC603493, KC603495-KC603497, KC603500, KC603501, KC603505-KC603510, KC603512, KC603527
Exocoetidae	NI	NI	3	KC603498, KC603502, KC603503
Halosauridae	*Halosaurus*	NI	1	KC603516
Molidae	*Ranzania*	*Ranzania laevis*	2	KC603525, KC603526
Myctophidae	*Lampadena*	*Lampadena atlantica*	1	KC603522
Myctophidae	*Diaphus*	NI	1	KC603519
Neoscopelidae	*Neoscopelus*	*Neoscopelus macrolepidotus*	1	KC603513
Scombridae	*Katsuwonus*	*Katsuwonus pelamis*	2	KC603511, KC603528
Sparidae	*Boops*	*Boops boops*	2	KC603531, KC603534
Sternoptychidae	*Argyropelecus*	NI	1	KC603514
Synaphobranchidae	NI	NI	2	KC603520, KC603524
Trichiuridae	NI	NI	2	KC603518, KC603523