Diet analysis of Kittiwake and Shag using DNA metabarcoding of faeces

Nina J O'Hanlon[1]*, David C Jardine[2], Jim Lennon[3], Dasha Svobodova[4], Robert L Swann[5], Robin M Ward[6], Elizabeth M Humphreys[1] and Barbara J Morrissey[4]

[1] BTO Scotland, Beta Centre (Unit 15), Stirling University Innovation Park, Stirling, FK9 4NF; [2] Hazel Cottage, 7 Barrmor View, Kilmartin, Lochgilphead, Argyll, PA31 8UN; [3] Shiants Seabird Research Group; [4] Institute for Biodiversity and Freshwater Conservation, UHI Inverness, 1 Inverness Campus, Inverness IV2 5NA; [5] Highland Ringing Group, 14 St Vincent Road, Tain, Ross-shire IV19 1JR; [6]Treshnish Isles Auk Ringing Group.

* Corresponding author: nina.ohanlon@bto.org

NOH ORCID: 0000-0001-6396-4518; EMH ORCID: 0000-0002-2570-400X; BJM ORCID: 0000-0002-7901-7678

Supplementary Material

Bioinformatics details

Sequencing data was automatically demultiplexed to separate forward and reverse fastq files per library using the onboard Illumina MiSeq Reporter software. Library sequence reads were further demultiplexed to sample using a custom Python script. Tapirs, a reproducible workflow for the analysis of DNA metabarcoding data (https://github.com/EvoHull/Tapirs), was used for taxonomic assignment of demultiplexed sequencing reads.

Raw reads were quality trimmed from the tail with a 5 bp sliding window (qualifying phred score of Q30 and an average window phred score of Q30) using fastp (Chen *et al.* 2018), allowing no more than 40% of the final trimmed read bases to be below Q30. Primers were removed by trimming the first 21 and 27 bp from the forward and reverse reads for 12S, respectively, and the first 26 bp from both reads for COI. Reads were then tail cropped to a maximum length of 170 bp and reads shorter than 90 bp were discarded for 12s, for COI reads were cropped to a maximum length of 313 bp and reads shorter than 120 bp were discarded.

Sequence read pairs were merged into single reads using fastp, provided there was a minimum overlap of 20 bp, no more than 5% mismatches and no more than 5 mismatched bases between pairs. Only forward reads were kept from read pairs that failed to be merged. For 12s a final length filter removed any reads longer than 190 bp to ensure sequence lengths approximated the expected fragment size (~170 bp). This was changed to remove reads longer than 330 bp for COI as the expected fragment size was 313 bp.

Redundant sequences were removed by clustering at 100% read identity and length (--derep_fulllength) in VSEARCH (Rognes *et al.* 2016). Clusters represented by less than three sequences were omitted from further processing. Reads were further clustered (--cluster_unoise) to

remove redundancies due to sequencing errors (retaining all cluster sizes). Retained sequences were screened for chimeric sequences with VSEARCH (--uchime3_denovo).

The final clustered, non-redundant query sequences were then compared against a curated UK marine fish reference database, Meta-Fish Lib (Collins *et al.* 2021) for 12s and against a curated marine eukaryote reference database, MARES (Arranz *et al.* 2020) for the COI data using BLAST (Zhang *et al.* 2000). Taxonomic identity was assigned using a custom majority lowest common ancestor (MLCA) approach based on the top 2% query BLAST hit bit-scores, with at least 90% query coverage and a minimum identity of 98% for 12s and 90% for COI. Of these filtered hits, 80% of unique taxonomic lineages therein had to agree at descending taxonomic rank (domain, phylum, class, order, family, genus, species) for it to be assigned a taxonomic identity. If a query had a single BLAST hit it was assigned directly to this taxon only if it met all MLCA criteria. Read counts assigned to each taxonomic identity were calculated from query cluster sizes. Lowest taxonomic rank was to species and assignments higher than order were classed as unassigned.

Table A1. Frequency of occurrence (% FO) of prey taxa identified in the diet of Kittiwake and Shag from Canna in 2023 from faeces and conventional methods (pellets and regurgitates; Swann et al. 2024). n = the sample size collected from each location. Taxa groups are shaded in grey to emphasise that the taxonomic resolution of prey was typically greater in the faeces than regurgitates/pellets.

			Kittiwake		Shag	
Taxa		Common name	Faeces (n=17)	Conventional (n=19[1])	Faeces (n=22)	Conventional (n=32[2])
Fish prey	Ammodytidae	Sandeel species	11 (65%)	17 (89%)	19 (86%)	28 (88%)
	Ammodytidae	Smooth Sandeel	-	-	6 (28%)	-
	Callionymidae	Dragonet species	-	-	-	9 (28%)
	Callionymidae	Common Dragonet	-	-	3 (14%)	-
	Cottidae	Bullrout species	-	-	-	1 (3%)
	Clupeidae	Herrings and sprats	6 (35%)	15 (79%)	-	1 (3%)
	Clupeidae	Atlantic Herring	13 (77%)	-	2 (9%)	-
	Clupeidae	European Sprat	13 (77%)	-	2 (9%)	-
	Gadidae	Cod species	3 (18%)	2 (11%)	15 (68%)	33 (94%)
	Gadidae	Atlantic Cod	-	-	1 (5%)	-
	Gadidae	Blue Whiting	1 (6%)	-	-	-
	Gadidae	Norway Pout	5 (29%)	-	1 (5%)	-
	Gadidae	Poor Cod	3 (18%)	-	5 (23%)	-
	Gobiidae	Goby species	-	-	-	1 (3%)
	Labridae	Wrasse species	-	-	-	8 (25%)
	Labridae	Goldsinny Wrasse	-	-	2 (9%)	-
	Lotidae	Rockling species	-	-	-	4 (13%)
	Pholidae	Butterfish species	-	-	-	1 (3%)
	Pleuronectidae	Flatfish species	-	-	-	2 (6%)
	Pleuronectidae	Lemon Sole	1 (6%)	-	-	-
	Pleuronectidae	Dab species	-	-	1 (5%)	-
	Scombridae	Atlantic Mackerel	1 (6%)	-	2 (9%)	-
	Triglidae	Gurnard species	-	-	1 (5%)	-
Marine invertebrate prey	Sepiolida	Bobtail Squid	-	-	-	-
	Crustacea	-	2 (12%)	-	-	-
	Crustacea	Shrimp species	-	-	-	1 (3%)
	Cephalopoda	Cephlapod species	-	-	-	3 (9%)
	Echinidae	Sea urchin species	-	-	-	2 (6%)
	Goneplacidae	Angular Crab	1 (6%)	-	-	-
	Mollusca	Mollusc species	-	-	-	10 (31%)
	Paguroidea	Hermit Crab species	-	-	-	1 (3%)
	Pilumnidae	Bristly Crab	1 (6%)	-	-	-
	Polybiidae	Marbled Swimming Crab	1 (6%)	-	-	-

[1] Regurgitate samples from chicks and adults, reflecting chick diet. [2] Pellets from 29 adults and regurgitates from three chicks.

References

Arranz, V., Pearman, W.S., Aguirre, J.D. & Liggins, L. 2020. MARES, a replicable pipeline and curated reference database for marine eukaryote metabarcoding. *Scientific Data* 7: 209.

Collins, R.A., Trauzzi, G., Maltby, K.M., Gibson, T.I., Ratcliffe, F.C., Hallam, J., Rainbird, S., Maclaine, J., Henderson, P.A., Sims, D.W. & Mariani, S. 2021. Meta-fish-lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes. *Journal of Fish Biology* 99:1446-1454.

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4: 2584.

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* 7: 203–14.