MOLECULAR ECOLOGY
Molecular Ecology (2017) doi: 10.1111/mec.13996

Genetic structure in the European endemic seabird, *Phalacrocorax aristotelis*, shaped by a complex interaction of historical and contemporary, physical and nonphysical drivers

EVANTHIA THANOU,*† STEFANO SPONZA,‡ EMILY J. NELSON,†§ ANNIKA PERRY,† SARAH WANLESS,† FRANCIS DAUNT† and STEPHEN CAVERS†
*Section of Animal Biology, Department of Biology, University of Patras, Patras GR-26504, Greece, †Centre for Ecology & Hydrology, Penicuik EH26 0QB, UK, ‡Department of Mathematics and Geosciences, University of Trieste, I-34127 Trieste, Italy, §Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK

Abstract

Geographically separated populations tend to be less connected by gene flow, as a result of physical or nonphysical barriers preventing dispersal, and this can lead to genetic structure. In this context, highly mobile organisms such as seabirds are interesting because the small effect of physical barriers means nonphysical ones may be relatively more important. Here, we use microsatellite and mitochondrial data to explore the genetic structure and phylogeography of Atlantic and Mediterranean populations of a European endemic seabird, the European shag, *Phalacrocorax aristotelis*, and identify the primary drivers of their diversification. Analyses of mitochondrial markers revealed three phylogenetic lineages grouping the North Atlantic, Spanish/Corsican and eastern Mediterranean populations, apparently arising from fragmentation during the Pleistocene followed by range expansion. These traces of historical fragmentation were also evident in the genetic structure estimated by microsatellite markers, despite significant contemporary gene flow among adjacent populations. Stronger genetic structure, probably promoted by landscape, philopatry and local adaptation, was found among distant populations and those separated by physical and ecological barriers. This study highlights the enduring effect of Pleistocene climatic changes on shag populations, especially within the Mediterranean Basin, and suggests a role for cryptic northern refugia, as well as known southern refugia, on the genetic structure of European seabirds. Finally, it outlines how contemporary ecological barriers and behavioural traits may maintain population divergence, despite long-distance dispersal triggered by extreme environmental conditions (e.g. population crashes).

Keywords: demography, European seabird, genetic structure, molecular clock, phylogeography, Pleistocene refugia

Received 9 December 2015; revision accepted 21 December 2016

Introduction

Studies of population genetics evaluate how genetic variation is partitioned across a species' range. Often, a species does not exist as a single panmictic population, but rather as several subdivided populations, each experiencing the effects of independent genetic drift and local adaptation to the various geographical locations (Avise 2000). Geographically separated populations may show a low rate or complete lack of immigration by individuals, and so reduced gene flow can lead to population differentiation, the first stage of speciation (Avise 2000). The geographical separation of

Correspondence: Evanthia Thanou, Fax: +44 (0) 131 4453943;
E-mails: ethanou@upatras.gr, thanouevanthia@gmail.com

populations may be a result of physical or nonphysical barriers that prevent dispersal and restrict gene flow, or simply of long distances between the ranges of each population (isolation by distance). The vagility of organisms like birds allows them to overcome many of these barriers. Seabirds in particular can disperse long distances, so high levels of gene flow might be expected among populations. Despite this, many seabird species exhibit substantial genetic structure in terms of significant divergence among breeding populations (Friesen *et al.* 2007). Thus, seabirds provide good models for studying speciation because physical barriers may have a small effect on their diversification, so the role of nonphysical ones may be more evident in seabirds than in less mobile organisms (Friesen 2015).

Physical barriers shaping the genetic structuring of organisms may be either historical (e.g. advance of ice sheets) or contemporary (e.g. land masses, ice or open sea for inshore species) (Avise 2000; Friesen *et al.* 2007). Many studies suggest that closely related species and subspecies of seabirds began to diverge in the Pleistocene, when distributional changes associated with glacial cycles may have induced population fragmentation and subsequent differentiation as a result of allopatric divergence in multiple glacial refugia (e.g. Avise *et al.* 2000; Moum & Árnason 2001; Morris-Pocock *et al.* 2008, 2010). Moreover, continental land masses can isolate populations in different ocean basins; for example, the presence of the narrow Isthmus of Panama may serve as an effective physical barrier to gene flow for several tropical seabirds (Avise *et al.* 2000; Steeves *et al.* 2005).

In the absence of an obvious physical barrier, other, nonphysical barriers must be responsible for the existence of genetic structure between populations of seabirds (Friesen 2015). Factors related to species life histories can act as behavioural barriers, which sufficiently prevent gene flow and promote genetic differentiation in birds (Avise 2000; Newton 2003). Species with narrow foraging ranges might be more likely to be genetically structured, as a consequence of lower rates of gene flow. Alternatively, population-specific nonbreeding or foraging distribution may be important in reducing contact between individuals from different colonies. Finally, strong philopatry is expected to lead to genetic differentiation among breeding populations through drift and⁄or selection (e.g. Friesen *et al.* 2007; Morris-Pocock *et al.* 2008).

Although several studies have been published to date (reviewed in Friesen 2015), approximately 1/3 of seabird species lack any population genetic analysis. For example, only eight of the approximately 40 extant Phalacrocoracidae species have been studied (Marion & Le Gentil 2006; Duffie *et al.* 2009; Barlow *et al.* 2011; Mercer *et al.* 2013; Calderón *et al.* 2014; Rawlence *et al.* 2014). Moreover, few of the published phylogeographical analyses concern the European avifauna and even fewer the seabirds of the Mediterranean region (e.g. Gómez-Diaz *et al.* 2006; Gay *et al.* 2007). Comparative phylogeographical studies, mostly of terrestrial plants and animals (e.g. Hewitt 2000), suggest that the palaeogeographical history of this region has shaped European biodiversity. Particularly during glacial periods, fragmented populations survived in southern refugia (Balkan, Italian and Iberian Peninsulas) and expanded northward in interglacial periods (Avise 2000; Hewitt 2000), although cryptic northern refugia have also been reported (Stewart *et al.* 2010). Sea level oscillations significantly altered the Atlantic and Mediterranean coasts and revealed land bridges within the Mediterranean (Taberlet *et al.* 1998; Bonfiglio *et al.* 2002), facilitating or limiting the dispersal of terrestrial and marine organisms, respectively. The Atlantic Ocean and the Mediterranean Sea were completely separated during the Messinian salinity crisis (MSC), until the tectonic collapse of the Gibraltar Strait (5.3 My ago) (Krijgsman *et al.* 1999). Nowadays, only this narrow channel allows passage between them, retaining their distinct oceanographic characteristics (Coll *et al.* 2010) and seabird communities' diversity (Zotier *et al.* 1999). As seabirds are usually associated with specific marine habitat features (Zotier *et al.* 1999), the Gibraltar Strait might represent both a historical physical barrier and a nonphysical ecological one.

Seabirds thus represent a useful system for assessing the importance of physical and nonphysical barriers, investigating the role of European refugia and specifically of the ecological and historical interplay between the Atlantic and the Mediterranean on phylogeographical patterns of marine species. Here, we focus on a Palaearctic endemic seabird, the European shag (Aves: Suliformes, Phalacrocoracidae), *Phalacrocorax aristotelis*. Within the recently published phylogeny of cormorants, Kennedy & Spencer (2014) suggest that this might be a monotypic genus (*Gulosus*). Currently, the species comprises three morphologically distinct subspecies: *P. a. aristotelis* (hereafter Atlantic shag), with a breeding distribution from northern Russia to the Atlantic coast of Iberia, *P. a. desmarestii* (Mediterranean shag), which breeds within the Mediterranean, and *P. a. riggenbachi* (African shag) distributed along the northern and western African coasts. So far, there is no available fossil evidence of the historical distribution of the shag (see Mlíkovský 2009) and no published studies regarding the putative biogeographical events that could have shaped its present distribution and genetic structuring.

Ecological or behavioural barriers might also reduce gene flow among shags, as they show high breeding philopatry (Aebischer 1995; Velando & Freire 2002) and

wide stretches of open sea probably impose a physical barrier to their movements (Wanless & Harris 2004). On the other hand, immature shags make occasional long-distance movements, up to several hundred kilometres (Wanless & Harris 2004), and adult shags also move between locations during nonbreeding periods (Sponza *et al*. 2013; Grist *et al*. 2014). However, it remains unknown to what extent such movements translate into effective gene flow.

Previous genetic analysis of this species (Barlow *et al*. 2011) focused mainly on the genetic structure of the Atlantic shag, although including one Mediterranean population (Corsica). Weak genetic structure was found among Atlantic populations, suggesting the existence of historical and present gene flow, while analyses failed to distinguish between the Atlantic shag and the Mediterranean shag. The authors hypothesized that the Corsican population could represent the extreme end of a cline in variation between *P. a. aristotelis* to the west and populations of *P. a. desmarestii* to the east and that the Atlantic colonies could have been colonized from southern Mediterranean refugia.

In this study, we sampled widely throughout the Mediterranean and collected new sequence data to develop a robust phylogeographical framework for the shag. We quantified population genetic structure, estimated gene flow rates and assessed phylogeographical patterns, timing of lineage divergence and historical demography, using recently developed coalescent models. We aimed to evaluate the effect of (i) historical events, such as the MSC or Pleistocene glacial cycles on intraspecific diversification and (ii) contemporary barriers, physical or otherwise (e.g. land masses and long distance between colonies) on the genetic structure of shag colonies. In the first case, we expected timing of intraspecific divergence points to mirror that of known historical events, helping identify the number and location of past refugia and reconstruct (re)colonization patterns. In the second case, we expected populations to show isolation-by-distance patterns and restricted contemporary gene flow between distant colonies.

Materials and methods

Study area and sampling strategy

We analysed a total of 519 specimens from 28 shag colonies (Table 1, Fig. 1A). The Mediterranean populations were represented with 72 samples (oral swabs, feathers, eggshells, blood and tissue) collected during 2006–2012 from eight shag colonies in the Aegean and Adriatic seas, which were pooled with samples from 20 colonies from the Atlantic and Corsica (Barlow *et al*. 2011). Samples were stored dry at −20 °C or in 95% ethanol at −80 °C.

DNA isolation, mtDNA sequencing and microsatellite genotyping

We extracted total DNA from the Aegean and Adriatic samples with the DNeasy Blood and Tissue Extraction Kit (QIAGEN), following the manufacturer's procedure.

For the mitochondrial DNA (mtDNA) analyses, we amplified a segment of the NADH dehydrogenase two (ND2) protein-coding gene and part of the noncoding control region one (CR1). We amplified the ND2 marker with primers and conditions as in Barlow *et al*. (2011), while for the amplification of CR1, we used several combinations of primers used in other Phalacrocoracidae species or specifically designed for *P. aristotelis* (for detail on amplification procedures, see Table S1, Supporting information). We then excised the amplification products from 1% agarose gel and purified them with the NucleoSpin Extract II DNA purification kit (MACHEREY–NAGEL). Both strands of the mtDNA segments were sequenced on an ABI Prism 3100 capillary sequencer (CEMIA, Larissa, Greece) for the Aegean and Adriatic samples, and at the Natural Environment Research Council (NERC) Biomolecular Analysis Facility (NBAF) at the University of Edinburgh for the Atlantic and Corsican samples.

We used the same amplification conditions and primers as in Barlow *et al*. (2011), to genotype the Aegean and Adriatic samples for seven microsatellite markers, that is three dinucleotide (Phaari02, Phaari05 and Phaari06) and four tetranucleotide (Phaari08, Phaari11, Phaari12 and Phaari16) motifs. We genotyped the products on an 8-capillary ABI 3500 Genetic Analyser (Applied Biosystems) and scored them in comparison with a size standard (GeneScan 600LIZ, Applied Biosystems), using the GeneMapper Software version 4 (Life Technologies/Applied Biosystems). As previously published genotypes for the Atlantic and Corsican shags resulted from fragments visualized via 6% denaturing polyacrylamide gels on a LI-COR 4200 IR2 genotyper (Barlow *et al*. 2011), we regenotyped 10 randomly chosen samples from the Atlantic and Corsica alongside the Aegean and Adriatic samples and cross-checked the results to ensure consistency.

Mitochondrial markers

Phylogenetic analyses. We used ClustalX version 2.0 (Larkin *et al*. 2007) to align the sequences for each mtDNA segment with homologous sequences from other phalacrocoracid and suliform species as outgroups for alignment and tree rooting (Table S4,

Table 1 Details on the *Phalacrocorax aristotelis* sampling localities and genetic diversity per colony. Genetic diversity values represent the number of individuals (*n*) analysed in each case, the nucleotidic diversity (*P*) and haplotype diversity (Hd), based on the concatenated mtDNA sequences of ND2 and CR1 markers, and mean number of alleles (Na), number of private alleles (Pa), allelic richness (Ra), averaged values of observed heterozygosity (H_o) and expected heterozygosity (H_e), and inbreeding coefficient (F_{IS}; significant values are bold, P < 0.05). based on the microsatellite data. MtDNA and microsatellite diversity values are given in populations with $n \geq 2$ and $n \geq 5$, respectively

					MtDNA diversity			Microsatellite diversity						
Geographical location	Abbr.*	Subspecies	Latitude	Longitude	*n*	*P* (%)	Hd	*n*	Na	Pa	Ra	H_o	H_e	F_{IS}
Aegean Sea					34			54						
Xironisi, Greece	XR	*desmarestii*	40°52′N	24°20′E	18	0.26	0.88	32	7.71	1	3.35	0.64	0.67	**0.058**
Sporades, Greece	SP	*desmarestii*	39°04′N	24°06′E	3	0.28	1.00	4	–	–	–	–	–	–
Evvoia, Greece	EV	*desmarestii*	37°59′N	24°15′E	8	0.30	0.96	9	6.29	1	3.62	0.69	0.66	0.011
Fournoi Isl., Greece	FO	*desmarestii*	36°28′N	37°17′E	1	–	–	2	–	–	–	–	–	–
Tilos Isl., Greece	TL	*desmarestii*	36°28′N	27°16′E	1	–	–	2	–	–	–	–	–	–
Crete Isl., Greece	CR	*desmarestii*	34°49′N	24°05′E	3	0.39	1.00	5	4.29	0	3.61	0.62	0.68	0.210
Adriatic Sea					7			18						
Oruda, Croatia	OR	*desmarestii*	44°32′N	14°35′E	4	0.08	0.50	9	4.43	1	3.12	0.74	0.60	−0.171
Brijuni, Croatia	BJ	*desmarestii*	44°54′N	13°45′E	3	0	0	9	5.57	0	3.45	0.70	0.64	−0.035
Corsica, France	CS	*desmarestii*	42°22′N	08°32′E	7	0.33	1.00	20	6.86	2	3.34	0.64	0.64	0.031
Spain					8			95						
Galicia, Spain	GC	*aristotelis*	42°13N	08°54′W	2	–	–	56	8.43	2	3.43	0.74	0.69	**−0.067**
Asturias, Spain	AS	*aristotelis*	43°33′N	06°59′W	4	0.38	1.00	24	9.00	0	3.75	0.73	0.74	0.026
Vizcaya, Spain	VZ	*aristotelis*	43°26′N	02°56′W	2	–	–	15	6.71	0	3.34	0.67	0.66	0.032
North Atlantic					23			332						
Île de Béniguet, France	IB	*aristotelis*	48°50′N	03°01′W	3	0	0	10	4.71	0	3.07	0.76	0.60	**–0.202**
Lambay, Ireland	LB	*aristotelis*	53°29′N	06°01′W	2	–	–	15	4.71	0	2.68	0.51	0.51	0.029
Ireland's Eye Ireland	IE	*aristotelis*	53°24′N	06°03′W	0	–	–	6	3.86	1	2.67	0.46	0.46	0.085
Staple Island, England	ST	*aristotelis*	55°37′N	01°37′W	0	–	–	41	9.57	1	3.69	0.65	0.72	0.060
Isle of May, Scotland	IM	*aristotelis*	56°11′N	02°33′W	8	0.09	0.46	27	8.29	1	3.31	0.65	0.65	0.023
Bullers of Buchan, Scotland	BB	*aristotelis*	57°25′N	01°49′W	0	–	–	51	8.57	0	3.14	0.60	0.61	0.026
Badbea, Scotland	BD	*aristotelis*	58°90′N	03°33′W	0	–	–	44	8.14	0	3.33	0.76	0.70	**–0.083**
Canna, Scotland	CA	*aristotelis*	57°30′N	06°33′W	0	–	–	16	6.86	0	3.16	0.64	0.61	–0.017
Skúvoy, Faroe Isl.	SV	*aristotelis*	61°46′N	06°49′W	1	–	–	5	3.57	0	2.88	0.60	0.53	**–0.012**
Flatey, Iceland	FT	*aristotelis*	65°22′N	22°54′W	2	–	–	20	5.71	0	3.10	0.62	0.62	0.029
Kjøer, Norway	KJ	*aristotelis*	58°53′N	05°26′E	2	–	–	18	5.43	0	2.84	0.55	0.55	0.038
Melstein, Norway	MS	*aristotelis*	63°57′N	09°32′E	0	–	–	13	5.71	0	3.42	0.62	0.69	0.155
Sklinna, Norway	SK	*aristotelis*	65°12′N	11°00′E	0	–	–	15	5.71	0	3.32	0.62	0.67	0.102
Røst, Norway	RS	*aristotelis*	67°26′N	11°54′E	2	–	–	28	7.57	1	2.93	0.52	0.55	0.069
Anda, Norway	AD	*aristotelis*	69°03′N	15°10′E	0	–	–	9	4.86	0	3.15	0.65	0.62	0.014
Hornøya, Norway	HN	*aristotelis*	70°23′N	31°08′E	3	0.04	0.30	14	5.14	0	2.90	0.51	0.53	0.001
Grand Total/Average					79	0.006	0.946	519	6.31	–	3.62	0.64	0.62	–

*Abbreviations stand for population names shown in Fig. 1A.

Supporting information). We performed final phylogenetic analyses on the mtDNA concatenated data set.

We computed within-group uncorrected genetic distances (*pi* and standard errors) and among groups *pi* for populations grouped according to their geographical origin (e.g. Aegean and Adriatic) and their respective phylogroup, using MEGA version 6 (Tamura *et al.* 2013).

We built phylogenies using Bayesian inference (BI) and maximum likelihood (ML). We used jMODELTEST version 2.1.4 (Darriba *et al.* 2012) to select the most suitable model of DNA substitution (Table S2, Supporting information). We tested three different partitioning strategies: (a) a single model of evolution for the entire mtDNA segment, (b) two separate evolution models and parameters for the two partitions (CR1 and ND2) and (c) four separate models of evolution, one for CR1 and three for the codon positions of ND2. We performed preliminary Bayesian analyses in MRBAYES version 3.2.3

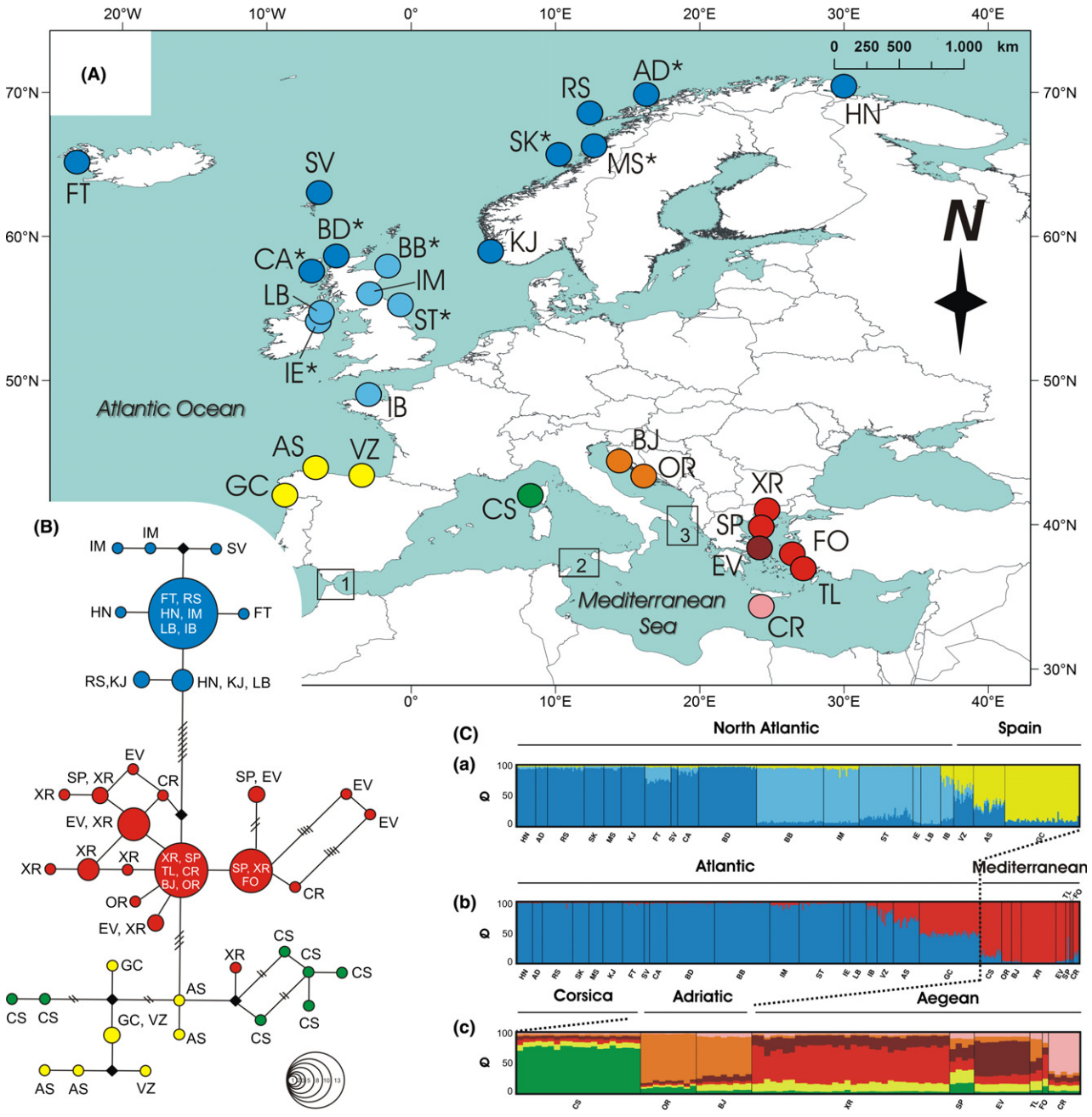

Fig. 1 Genetic structure of the European shag. (A) Sampling localities: locality abbreviations and references are defined in Table 1. Asterisks mark populations that were analysed for microsatellites only. The location of possible barriers to gene flow discussed in the text is also shown 1: Gibraltar Strait 2: Siculo-Tunisian Strait and 3: Otranto sill. Colours reflect the genetic clustering shown in (B) and (C). (B) Reduced median-joining network of mtDNA haplotypes obtained for the respective populations. Hash marks crossing line connections represent mutational steps. Haplotype frequencies are proportional to the size of the circle, and the actual number of individuals per haplotype is shown in the embedded standard. (C) Microsatellite loci bar plot obtained with STRUCTURE for (a) Atlantic populations only, (b) all studied populations and (c) Mediterranean populations only. *Q* values indicate membership probability of an individual to a given cluster.

(Ronquist & Huelsenbeck 2003; Ronquist *et al.* 2012) and used Bayes factors (Table S2, Supporting information) to choose the most appropriate partitioning strategy by calculating the marginal likelihood for all analyses with TRACER version 1.5.0 (Rambaut & Drummond 2009). The best partitioning strategy for the final analysis was (c).

We performed Bayesian analysis in MRBAYES with the default heating values for eight incrementally heated

Markov chains of $3*10^6$ generations and sampling intervals of 100 generations. Substitution model parameters were unlinked across partitions, and each subset was allowed to have its own rate. We run four independent analyses so that global likelihood scores, individual parameter values, topology and nodal support could be compared to check for local optima. We used TRACER to confirm stationarity (all parameters with an ESS > 200) and decide on the number of 'burn-in' trees (25%). All the remaining (post-burn-in) trees were used to estimate the topology of a 50% majority rule tree and its posterior nodal probabilities (p*P*).

We carried out partitioned ML analysis with RAXML 7.5.2 (Stamatakis 2006), with each partition having its own GTR-GAMMA model. Nodal support of the trees was tested via 1000 bootstrap replicates (bp).

Estimation of divergence times. We estimated the time of the most recent common ancestor (TMRCA) for nodes of interest observed in the mtDNA gene tree with the package BEAST version 1.8.0 (Drummond *et al.* 2012). To calibrate the molecular clock, we used the split between *P. aristotelis* (*Gulosus aristotelis*) and *Nannopterum* + *Leucocarbo* (Kennedy & Spencer 2014). To account for possible errors due to the use of a secondary calibration point, we applied a normal distribution with a mean value and standard deviation of 10 ± 0.5 My (setting the 97.5 intervals of the distribution at 9.0–11.2 Mya, as inferred by the primary analysis of Kennedy & Spencer 2014). Such an ancient calibration might underestimate substitution rates and therefore overestimate divergence times. To avoid that, we also applied a second prior on the substitution rates for the ND2 and CR1 using the ranges of published rates for Suliformes and other avian species (for detail, see Table S3, Supporting information). Specifically, we allowed uniform distributions for both priors with maximum and minimum values of substitutions/site/My of 0.0048–0.0090 and 0.025–0.1 for ND2 and CR1, respectively. Analyses were run under a relaxed molecular clock (uncorrelated lognormal), with a 'Coalescent: Constant Size' prior on rates of cladogenesis. We performed four independent runs, each with a chain length of $2*10^7$ iterations and trees sampled every 1000 generations. We checked for convergence with TRACER to (ESS values > 200), and after a burn-in of 10%, runs were combined in LogCombiner and TreeAnnotator from the BEAST package and the outcome was visualized with FIGTREE version 1.4.2 (Rambaut 2014).

Estimation of dispersal/vicariant events. To better understand the biogeographical history of the species, and to reconstruct ancestral distribution areas, we used a statistical approach of the dispersal–vicariance analysis (S-DIVA; Yu *et al.* 2010) as implemented in the program RASP version 3 (Yu *et al.* 2015). This approach accounts for uncertainty in the phylogenetic estimate, producing a probability *P* of the proposed ancestral range at each node. We used the mtDNA tree produced with BEAST and assigned the terminal nodes to geographical areas, based on their present distribution. We used an input of 10 000 trees produced with BEAST to estimate the probability and run the analysis with default 'Optimize' settings and under the 'Allow Reconstructions' option.

Network and demographic analyses. We estimated haplotype (Hd) and nucleotide diversity (*P*) values, number of haplotypes (h) and segregating sites (S) found in the concatenated mtDNA data set with DNASP version 5 (Librado & Rozas 2009). We constructed a median-joining (MJ) network of haplotypes, using NETWORK 4.6.1.2 (Bandelt *et al.* 1999).

We inferred the demographic fluctuations of the main identified phylogenetic units and for the total of our samples by two separate approaches. First, we tested for deviations from the neutral Wright–Fisher model consistent with a population expansion under the neutrality hypothesis. We calculated Fu's *Fs* and R_2 statistics in DNASP, based on segregating sites and assuming no recombination, and their respective significance, with 10 000 coalescent simulations and 0.95 confidence interval. Significantly negative *Fs* values (for large population sample sizes ~ 50) and significantly positive R_2 values (for small ones ~ 10) can be interpreted as signatures of population expansion (Ramos-Onsins & Rozas 2002).

Second, we constructed Bayesian skyline plots (BSPs) with BEAST, under the coalescent tree prior, with the piecewise constant model and number of groups set to half the sample size per group with a maximum of 10. We used the most suitable model for the entire mtDNA segment, a strict clock prior and a uniform distribution rate of 0.0055 substitutions/site/MY, which corresponds to the mean mutational rate estimated for our concatenated data set. We performed two runs, each with $8*10^7$ generations and log parameters sampled every 1000 generations. We assessed the ESS of parameters and generated BSPs with TRACER, after discarding a burn-in of 10%.

Microsatellite markers

Genetic variability and population differentiation. Due to low sample sizes, we excluded three Aegean populations (SP, TL and FO) from the following analyses. We checked genotypic data for amplification errors and for the presence of null alleles using MICROCHECKER version 2.2.3 (Van Oosterhout *et al.* 2004). We tested for linkage disequilibrium and Hardy–Weinberg equilibrium

(HWE) with GENEPOP version 4.2 on the online server (Rousset 2008), using an exact probability test (Markov chain parameters: 10 000 dememorizations, 100 batches, 1000 iterations per batch), with the Bonferroni's correction. In case of deviation from HWE, we also tested whether this was attributed to heterozygosity deficit or excess.

We calculated estimators of genetic variability, such as allele frequency, average number of alleles per locus (N_a), observed (H_o) and expected (H_e) heterozygosities for each microsatellite locus and each population with GENALEX 6.5 (Peakall & Smouse 2012). We estimated the inbreeding coefficient (F_{IS}) with randomization test for its significance and allelic richness (Ra) as a standardized measure of the number of alleles corrected by sample size using FSTAT version 2.9.3.2 (Goudet 2001). We also obtained the traditional estimators of genetic population differentiation, F_{ST} (Weir & Cockerham 1984) and R_{ST} (Slatkin 1995), as well as pairwise estimates of F_{ST} and R_{ST} between all populations with sample sizes ≥5, under the Bonferroni's correction, with FSTAT and SPAGEDI version 1.4 (Hardy & Vekemans 2002), respectively.

Testing for bottleneck and isolation effects. We tested the hypothesis of a bottleneck effect with BOTTLENECK version 1.2.02 (Cornuet & Luikart 1997). As the best-fit mutation model was unknown, we tested for both the SMM (stepwise-mutation model) and the two-phase model (TPM = 70% SMM, geometric distribution = 0.36). We evaluated the significance of heterozygosity excess using the Wilcoxon signed-rank test (10 000 simulation replicates) and the inspection of the allele frequency distribution curve; swift from a normal L-shaped distribution is an indicator which discriminates bottlenecked populations from stable ones (Piry *et al.* 1999).

We also tested for the presence of isolation by distance using Mantel's test (Mantel 1967) implemented in GENALEX between the matrix of linearized pairwise F_{ST} $[F_{ST}/(1 - F_{ST})]$ and a matrix of log-transformed geographical distance between populations estimated from geographical coordinates. We used two different approaches to construct the geographical matrixes: (i) using the modified Haversine formula algorithm which takes into account the spherical shape of the earth and (ii) calculating Euclidean distances in R (version 3.2.1, R Core Team 2014, packages 'sp', 'raster', 'rgdal', 'geosphere' and 'gdistance'), where distance between two colonies was taken following the coastline with cross-sea distances measured between land masses at the closest points. We estimated the significance for each test with 10 000 permutations.

Population clustering. We used the Bayesian clustering analysis implemented in STRUCTURE version 2.3.4 (Pritchard *et al.* 2000) to infer genetic structure for all the studied populations and samples. We tested the probability of different numbers (1–10) of independent genetic clusters (K) with five runs ($1*10^6$ iterations each, 50% burn-in), assuming correlated allele frequencies and using the admixture model with and without the LOCPRIOR setting, which takes into account sample location and is expected to perform better when genetic structure is weak or when the number of loci is small (< 20; Hubisz *et al.* 2009). The vector specifying the degree of admixture between each subpopulation (alpha) was inferred from our data, the distribution of allele frequencies (lambda) was estimated from our data ($\lambda = 0.62$ for $K = 1$), and the prior for F_{ST} was specified with a mean value and standard deviation of 0.01 ± 0.05.

For choosing the optimal number of clusters for our data, we estimated the log likelihood given the number of clusters [ln $P(X/K)$] (Pritchard *et al.* 2000) and the second-order rate of change of ln $P(X/K)$ (ΔK) (Evanno *et al.* 2005) over all five runs using STRUCTURE HARVESTER online Web server (Earl & Von Holdt 2012). As different solutions may result from replicated cluster analyses, we used CLUMPAK software (Kopelman *et al.* 2015) with the 'greedy' option and 10 000 random input orders to find the optimal individual alignments of replicated cluster analyses and to visualize the final results.

Results

Mitochondrial DNA phylogeography

We amplified 109 sequences of 578 bp for ND2 and 121 sequences of 624 bp for CR1, resulting in ten haplotypes (10 polymorphic sites) and 28 haplotypes (30 polymorphic sites), respectively. In the concatenated mtDNA data set, we included 79 sequences and identified 39 haplotypes (38 polymorphic sites) (Table S4, Supporting information). We found higher values of nucleotide diversity (P = 0.26–0.39%) in the Aegean, Corsican and Spanish populations, while the Adriatic and N. Atlantic populations showed lower values (0–0.09%) (Table 1). The genetic distances within and among major geographical groups were also relatively low (pi = 0–0.8%) (Table S5, Supporting information).

All phylogenetic analyses (gene trees) grouped populations in three strongly supported and distinct groups: the first phylogroup clustered all the N. Atlantic populations (Norway, Iceland, Faroe Is., England, Ireland, Scotland and France), the second the Spanish and Corsican populations and two individuals from the Aegean (XR) and the third all the remaining individuals from the Aegean and the Adriatic populations (Fig. 2). The phylogenetic relationships among the three

phylogroups were not fully resolved, but all southern populations (Spanish, Corsican, Adriatic and Aegean) were grouped in one clade with medium support (bootstrap $P < 90$, $pP < 70$). Each of the three phylogroups was further divided, although most subgroups showed medium-to-low support. The N. Atlantic populations (N. Atlantic phylogroup) were clustered in two subgroups (A1 and A2). Several subgroups were included within the Spanish/Corsican (S1, S2, C1 and C2) and the Aegean/Adriatic phylogroups (G1–G4). In some cases, these subgroups correspond to geographical regions; for example, the majority of the Corsican and most of the Adriatic shags were clustered in distinct clades (C1 and G3, respectively; Fig. 2).

According to our molecular dating (Fig. 2), the MRCA of all the European shags lived approximately 700 kya. The N. Atlantic phylogroup was then isolated from its sister clade of Spanish/Corsican + Aegean/Adriatic shags. The latter was further split approximately 500 kya, separating the western populations (Spain and Corsica) from the eastern Mediterranean ones (Adriatic and Aegean). Subsequent radiation occurred within each of the three phylogroups resulting in the divergence observed in current populations,

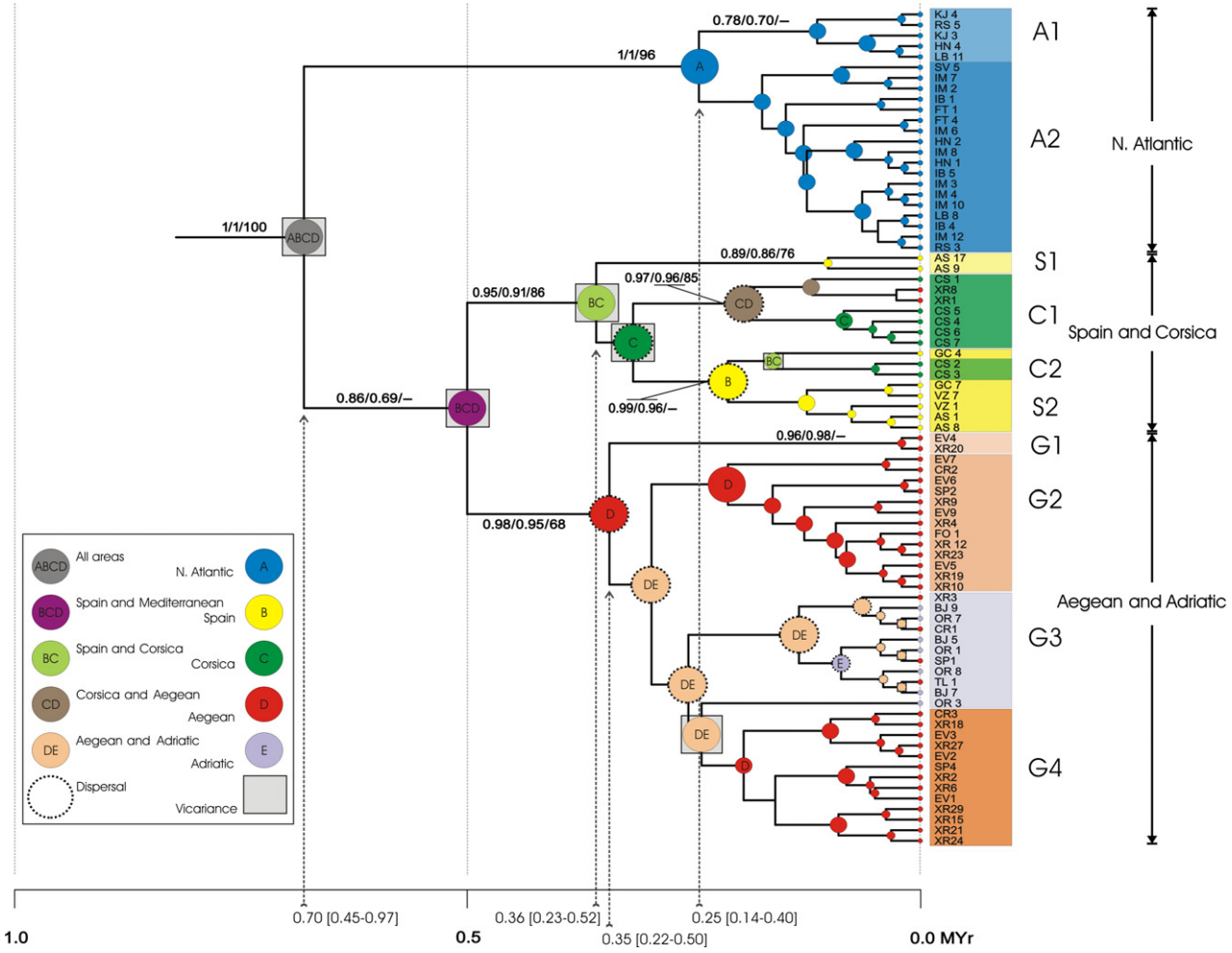

Fig. 2 The chronogram produced by BEAST, showing phylogenetic relationships and estimated divergence times of the European shag lineages. Letter codes with numbers in terminal nodes refer to mtDNA haplotypes of the concatenated data set (Table S4, Supporting information). Letters correspond to the respective population and locality shown in Fig. 1A. Statistical support is given above branches; numbers shown in order for the respective analysis (BEAST posterior probabilities/MrBAYES posterior probabilities/RAxML bootstrap values), only if $pP > 0.50$ and bootstrap values $P > 50$. Numbers at the axis on the bottom refer to MY since present; mean estimated times in MY and the 95% intervals are noted for each major node corresponding to the three phylogroups (N. Atlantic, Spain and Corsica, and Aegean and Adriatic) and their common ancestor. Subgroups within each major clade are shaded. The cladogram obtained with S-DIVA is also presented, imposed over the chronogram. On each node, circles and within letters correspond to the defined distribution area of the common ancestor (as explained in the inset) and the effect of vicariant/dispersal events are shown with dashed-line circles or shaded squares, respectively.

which was dated at approximately 350 kya for the Spanish/Corsican and the Aegean/Adriatic phylogroups and more recently at approximately 250 kya, for the N. Atlantic one.

The biogeographical history of the species seems to be a result of several vicariant and dispersal events and their combinations, according to the S-DIVA analysis (Fig. 2). Of the 14 divergence events found, six corresponded to vicariance, one to a vicariance/dispersal event and seven to dispersal events, all showing high values of probability ($P = 1.0$), with the exception of the nodes connecting subgroups S2/C2 and G3/G4 ($P = 0.31–0.56$). One inferred area of ancestral distribution was proposed for each of the studied nodes. The common ancestor of European shags was distributed from the N. Atlantic through Spain and as far as the eastern Mediterranean, covering all the areas of the species' present distribution. Vicariance has acted to split (a) N. Atlantic populations from all the remaining ones and (b) the latter populations into two groups: one in the eastern Mediterranean and one in the western Mediterranean and Spain. In the west, ancestral populations were originally distributed in Spain (AS) and Corsica, while the east Mediterranean ancestral populations were restricted in the Aegean. Several independent dispersal events were inferred from internal nodes, particularly throughout the Mediterranean Sea and between Corsica and Spain.

Mitochondrial DNA demographic analyses

The network reconstruction (Fig. 1B) retrieved the same relationships as in the phylogenetic analysis. Haplotypes were connected in three major clusters matching the three phylogroups, with the Aegean/Adriatic haplotypes placed in the central position. The N. Atlantic and the Aegean/Adriatic clusters showed a starlike topology with one or two common haplotypes in the middle, connected to several rare ones in the periphery. Most of the Spanish/Corsican haplotypes were rare or even unique to individual populations.

Fs values for each of the three phylogroups and for all samples were significantly negative and with strong statistical support ($P < 0.01$; Table 2). All R_2 values were positive, but none showed a statistical significance ($P > 0.05$; Table 2). Given the relatively small size of the samples tested for each phylogroup ($n = 17–39$), we expect the R_2 values to better represent their demographic history. Thus, although there is evidence of population expansion in all the phylogroups, it seems to be rather gradual than sudden. Moreover, it is probably recent, as shown from the estimated times for each phylogroup (Table 2). The skyline plots gave analogous results (Fig. 3). The Spanish/Corsican and the Aegean/Adriatic populations underwent a gradual and mild increase of their population sizes during the last 200 000 and 80 000 years, respectively, while the N. Atlantic populations had a constant population size during the last 30 000 years.

Genetic diversity based on microsatellite data

In total, we genotyped 519 European shags from 28 populations at seven polymorphic microsatellite loci (Table S6, Supporting information). Across all populations, the total number of alleles detected per locus ranged from two (Phaari05 in populations HN, AD, FT, SV, IE, LB, IB) to 17 (Phaari11 in ST). There was evidence for a low frequency of null alleles at loci Phaari02 (BB, CS, XR) and Phaari11 (XR), with possibility $P \leq 0.01$. In most populations, there was no evidence of departure from gametic equilibrium between any pair of loci ($P > 0.05$). Only two locus pairs in individual populations (Phaari06–Phaari12; IM and Phaari11–Phaari16; BD and IE) strongly deviated from equilibrium ($P < 0.001$), showing significant deficit or excess of heterozygotes; the former was found in SK, KJ, BB, ST, AS and XR for Phaari11, CA for Phaari16, BD for Phaari12 and AS for Phaari06, and the latter in BD and XR for Phaari05 and Phaari08, and in GC for Phaari02 ($P < 0.01$).

Based on the mean multilocus inbreeding coefficient per population (mean $F_{IS} = -0.202$ to 0.210), only XR, GC, IB, BD and SV returned significant departures from random mating. For XR, a deficit of heterozygotes was detected and a slight excess of heterozygosity was found in all other cases. There were consistently high levels of genetic diversity (mean $H_e = 0.46–0.74$) and consistently low levels of allelic richness (mean Ra = 2.67–3.75) for all locus–colony combinations. Private alleles were found in nine populations (Table 1).

The genetic differentiation between populations was moderate but revealed a statistically significant structure. The pairwise F_{ST} values ranged from 0 (for the population pair SK–MS) to 0.35 (CR–RS) with a global F_{ST} of 0.085 ($P \leq 0.001$, SE = ±0.043), and R_{ST} ranged from to 0 (for several pairs of populations within the North Atlantic, within the Aegean and between Aegean and Adriatic) to 0.57 (OR–KJ), with a global estimate of 0.23 ($P \leq 0.001$, SE = ±0.043) (Table S7, Supporting information).

Population structure based on microsatellite data

The results from cluster analysis with STRUCTURE converged to $K = 2$ in all independent runs (assuming correlated allele frequencies, using LOCPRIOR, and an estimated $\lambda = 0.6239$; mean ln $P = 11\ 959$/mean

Table 2 MtDNA diversity (for both ND2 and CR1 markers) and summary statistics (Fu's Fs and Ramos-Onsins and Rozas's R_2) of population size changes for each phylogroup presented in Fig. 2. Tests of significance are provided (10 000 coalescent simulations at 0.95 confidence interval) (significant values are bold, $P < 0.05$). Values of tau and t ($\tau = 2\mu t$, where μ is the mutational rate and t the time since expansion) are also provided, using the mutational rate $\mu = 0.0055$ substitutions/MY estimated in BEAST

	Aegean and Adriatic	Spain and Corsica	N. Atlantic	All samples
Number of sequences	39	17	23	79
Number of haplotypes	16	15	8	39
Number of polymorphic sites	15	15	7	38
Tau	1.492	4.779	1.062	3.780
t (years)	135 600	435 400	96 500	343 600
Fs	**−8.446**	**−9.264**	**−3.622**	**−19.393**
95% CI	−4.283 – 5.331	−4.232 – 5.120	−2.821 – 3.921	−7.687 – 0.080
R_2	0.077	0.150	0.086	0.086
95% CI	0.059 – 0.192	0.087 – 0.211	0.083 – 0.239	0.048 – 0.160

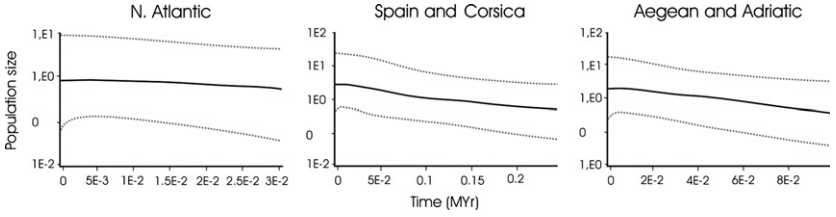

Fig. 3 Bayesian skyline plots based on the mtDNA data for each of the three phylogroups (N. Atlantic, Spain and Corsica, and Aegean and Adriatic), showing the median population size changes through time (solid lines), as well as the upper and lower 95% HPD (grey lines).

similarity score of five runs was 98%). According to the estimated values of ln $P(X/K)$, Evanno's ΔK and Q (membership probability of each individual to a given cluster), this scenario is strongly supported against all alternative clustering for $K = 3$–10 (Fig. S1a, Supporting information). The same outcome persisted when different options regarding the allele frequencies, admixture model and values of lambda were tested, although alternative options returned lower mean lnP values. These two clusters (Fig. 1Cb) grouped all Atlantic and all Mediterranean populations, respectively. Although clearly placed within the Atlantic cluster, the Spanish populations appear admixed, especially GC, which showed the lowest membership proportion ($Q = 0.52$).

To further investigate the population structure within each of the two groups, we repeated two independent analyses, one including only the Atlantic and the other including only the Mediterranean populations. The Atlantic populations (Fig. 1Ca) were assigned to three clusters (Fig. S1b, Supporting information) ($K = 3$, five runs with mean ln $P = -9\,407$/mean similarity score at 96%). The first clustered all Norwegian populations, SV, FT and the Scottish populations BD and CA, the second clustered the remaining N. Atlantic populations and IB (France), and the third clustered the Spanish ones; populations FT, IB, VZ and AS were admixed. For the Mediterranean populations (Fig. 1Cc), an ambiguous clustering was observed, resulting in either $K = 4$ or $K = 6$ (Fig. S1c, Supporting information), with the latter showing a slightly higher probability (mean ln $P = -2\,051$/mean similarity score at 85%). The ambiguity arose from different clustering of the Aegean populations, whereas Corsican and Adriatic populations formed two consistently distinct groups. Within the Aegean, at least three additional clusters can be defined for the XR, EV and CR populations.

The bottleneck hypothesis was rejected for most populations under either mutational model (normal L-shaped allele frequency distributions and no evidence of heterozygosity excess; $P > 0.01$). Under the SMM model, some of the N. Atlantic populations (HN, RS, KJ, CA, BB, IM, ST, IE), as well as VZ (Spain) and XR (Greece), showed a normal distribution despite some evidence of heterozygote deficiency ($0.05 \geq P > 0.01$). Only the TPM model revealed a shifted distribution and statistically significant evidence of heterozygosity excess ($P > 0.01$), suggesting past bottleneck in three populations, namely MS, GC and OR.

Genetic distances showed a significant positive correlation with geographical distance between populations only when distances were calculated through sea and along the coastlines ($R = 0.730$, $r^2 = 0.4922$, $P < 0.001$). Distances measured over land with the Haversine formula algorithm were not found to have a significant effect on genetic differentiation between populations ($R = -0.020$, $r^2 = 0.0004$, $P > 0.05$).

Discussion

Wider sampling and independent marker analyses revealed a more complex population genetic structure than previously suggested for the European shag (Barlow *et al.* 2011). This structure was mostly concordant for nuclear and mitochondrial markers; the two extremes of the species' distribution evaluated here, that is the North Atlantic and the eastern Mediterranean, are highly diverged. The placement of Spanish and Corsican populations was not conclusively resolved, although mtDNA gene trees clustered them in one phylogroup and favoured their closer relationship with the eastern Mediterranean. Spanish populations (and, to a lesser degree, the French population) were admixed, showing evidence of current gene flow from the north (N. Atlantic) and the east (mainly Corsica). These points are discussed below in the light of historical and present, physical and nonphysical barriers that drive the differentiation of a European marine organism.

Historical biogeography

The first hypothesis tested here was that intraspecific diversification of the shag has been driven primarily by historical events, either during the MSC (at approximately 5 Mya) or during the more recent Pleistocene climatic changes. Extant shag populations are clearly separated into two phylogenetic groups, one in the northern (N. Atlantic) and the other in the southern part of the species' current distribution (Mediterranean and possibly Spain). Our time estimates indicate that this split postdated the MSC and occurred more recently, during the Pleistocene. This is in agreement with the absence of fossil records for the Mediterranean Sea avifauna, suggesting that most seabird species did not survive the development and maintenance of a desiccated Mediterranean basin during the Late Miocene MSC (Mlíkovský 2009). According to the S-DIVA analysis and molecular dating (Fig. 2), the European shag had a wide past distribution throughout the Mediterranean and Atlantic which was later fragmented by a vicariant event that occurred between 1 Mya and 450 000 ya.

The Pleistocene epoch (2.6–0.11 Mya) was a time of extraordinary oscillations in global climate. The effect of these climatic changes on the geographical distributions of species was profound, and it has been shown that many of the phylogenetic separations that led to extant sister species of birds were initiated in the Pliocene (5.3–2.6 Mya) and completed in the Pleistocene (Klicka & Zink 1997), while many pronounced phylogeographical separations within avian species have occurred within the last 2 My (Avise 2000). These separations, also revealed in the shag phylogeny, were probably the result of extended ice masses driving isolation in combination with environmental changes on land and at sea (e.g. sea level fluctuations, vegetation, current circulation, productivity and food availability).

Pleistocene effects on the phylogeographical structure of conspecific avian populations have been demonstrated in seabirds across a wide range of environments, for example arctic: razorbill (*Alca torda*, Moum & Árnason 2001) and two murre species (*Uria aalge*, Morris-Pocock *et al.* 2008; *Uria lomvia*, Tigano *et al.* 2015); tropical: masked booby (*Sula dactylatra*, Steeves *et al.* 2005); and temperate: Cory's shearwater (*Calonectris diomedea*, Gómez-Diaz *et al.* 2006). In most of these cases, Atlantic populations were genetically differentiated from their conspecifics in other marine regions (Moum & Árnason 2001; Morris-Pocock *et al.* 2008). Additionally, a similar phylogeographical boundary between the Atlantic and the Mediterranean has been found among seabirds (e.g. the European storm petrel *Hydrobates pelagicus*, Cagnon *et al.* 2004; Cory's shearwater, Gómez-Diaz *et al.* 2006), showing congruent differentiation patterns with the shag and divergence times dating back to the Pleistocene.

In Europe, pronounced global cooling on a 100 000-year cycle resulted in continental ice sheets that extended as far south as the 52nd parallel (Berger 1984). The separation zone for the shag populations currently lies between 43° and 47°N (between France and Spain; Fig. 1A) or at approximately 50°N, excluding the admixed France colony (IB). Land masses (or ice masses in this case) can provide a substantial physical barrier to dispersal, as shags do not typically fly over land. The accumulation of ice resulted in lower sea level during the glacial periods, establishing land connections between Mediterranean islands and the mainland (e.g. Sicily and the southwestern tip of Italy, Bonfiglio *et al.* 2002) or coastlines (e.g. the northeastern and northwestern Adriatic coasts, Taberlet *et al.* 1998). This could have formed a physical barrier to dispersal within the Mediterranean and promoted isolation between the east and the west Mediterranean approximately 500 000 ya (Fig. 2).

The hypothesis of a postglacial colonization of the North Atlantic from southern refugia is not supported by our molecular evidence. The network analysis (Fig. 1B) suggests the existence of two originally isolated ancient stocks in the N. Atlantic and the eastern Mediterranean. The respective isolation of shag populations during glaciations and the inferred areas of ancestral distribution (S-DIVA; Fig. 2) imply the existence of both northern and southern refugia. The location of refugia in southern Europe has been well documented for several species of terrestrial animals with diverged populations originating from the Iberian, Italian or Balkan peninsulas (Avise 2000; Hewitt 2000; Stewart *et al.* 2010). This study provides further evidence for an

important role of these refugia in the divergence of seabirds. In contrast, the existence of cryptic refugia in northern Europe has only recently been proposed (Stewart *et al.* 2010). Studies of genetic variation in the herring gull complex (*Larus argentatus,* Liebers *et al.* 2004) and the thick-billed murre (Tigano *et al.* 2015) suggested the existence of a refugium in the Northeast Atlantic during the last glacial maximum (LGM), probably located on the islands of the Arctic Ocean. Interestingly, both in our study and in that of Tigano *et al.* (2015) the Hornøya population (HN) showed private haplotypes, suggesting that this may represent a refugial colony within the Northeast Atlantic. This region may have acted as a glacial refugium for many organisms, cold-adapted or with a high tolerance of cold conditions, due to the presence of ice-free corridors along the Atlantic coastal areas of north Scandinavia during most of the LGM (Parducci *et al.* 2012).

After the subdivision of the ancestral shag population, the three diverged phylogroups expanded almost simultaneously at approximately 350 000–250 000 ya (Fig. 2). In the case of the N. Atlantic, populations possibly expanded following the retreat of ice masses in a southwest direction, while Aegean populations colonized the Adriatic Sea. Multiple dispersal events were documented in the west Mediterranean and the Spanish coasts, showing evidence of movement through the Gibraltar Strait in both directions, as well as from Corsica to the eastern Mediterranean (Fig. 1B, 2). A general trend towards population stability within each of the three phylogroups has continued over the past 300 000 years, as shown by the BSP analyses (Fig. 3).

Influence of current physical and nonphysical barriers

Regardless of the detectable imprint that historical events have left on the population structure of the European shag, current physical and nonphysical barriers also act to shape the genetic diversity of the species. Microsatellite data (Fig. 1C) depict two genetic clusters that correspond to the populations distributed on either side of the Gibraltar Strait, with some admixture in the Spanish populations. This suggests a barrier restricting movements between Atlantic/Spanish and Mediterranean populations. The Gibraltar Strait may act as a major physical boundary between Atlantic and Mediterranean populations of seabirds and other marine organisms (see Gómez-Diaz *et al.* 2006; Quéré *et al.* 2012 and references therein). However, we cannot exclude the possibility of rare events of immigration between Corsica and Spain, given the genetically admixed Spanish populations and especially Galicia (GC), and the evidence provided by S-DIVA analysis (Fig. 2) that suggests past interglacial dispersal across the Gibraltar Strait.

Genetic structuring was also present within the Atlantic (Fig. 1Ca) and the Mediterranean (Fig. 1Cc). The Atlantic populations were divided into three clusters: one of them included the Spanish populations and the other two the North Atlantic ones in a northwestern–southeastern transition (Norway, Iceland and the northwestern colonies of the UK vs. southeastern UK and France). A similar population structuring among N. Atlantic colonies is weakly reflected in the mtDNA phylogeny (Fig. 1B, 2); therefore, it could partially be attributed to historic fragmentation of populations. As demonstrated, these populations probably originate from a small genetic stock; it is likely that northwestern and southeastern colonies are related to genetically distinct founder populations.

Within the Mediterranean, the observed genetic structure corresponds to at least four geographical regions, that is Corsica, Adriatic, Aegean and Crete. There is no evident physical barrier between them that might prevent overseas movements, but several hydrological boundaries have been proposed, that is underwater ridges that affect the prevailing current flows in the different Mediterranean basins, shaping their hydrological features (e.g. salinity and cold-water upwelling) (Coll *et al.* 2010). Examples of such barriers are the Siculo-Tunisian Strait (separating the west and the east Mediterranean) and the Otranto sill (isolating the Aegean/Ionian and Adriatic seas), which have promoted genetic structure within the last 30 000 years in several marine organisms (e.g. common cuttlefish *Sepia officinalis,* Pérez-Losada *et al.* 2007; sea bass *Dicentrarchus labrax*, Quéré *et al.* 2012; goby *Pomatoschistus tortonesei*, Merji *et al.* 2009). Therefore, it is possible that genetic structuring within Mediterranean shags is induced by ecological barriers such as local adaptation to prey movements.

Shag populations also display genetic structure within geographical regions (North Atlantic and Aegean) in the absence of any apparent physical barriers, suggesting a possible role for other factors such as (i) founder or bottleneck events that affect local population dynamics, (ii) philopatry and (iii) isolation by distance between colonies in respect of the species' flight capacity. Breeding populations of shags may exhibit periodic crashes followed by rapid population growth (Frederiksen *et al.* 2008). Although several crashes have been reported throughout the species range, the bottleneck hypothesis was rejected in most of the populations studied here, while both historical and contemporary demographic analyses (Tables 1, 2 and Fig. 3) suggest population stability. One explanation could be that crashes affect a given colony, but their overall impact on the shag population is minimal and/or cannot be detected with our approach. Another possible explanation could be that crashes are sometimes the result of

reproductive failure affecting the population size during the crash period but not the survival of adults who will probably breed in their natal colony during the following reproductive period (Potts *et al.* 1980; Aebischer 1986; Thanou 2013). Moreover, even after periodic large-scale mortality events, shag populations are known to recover rapidly (boom-and-bust population dynamics, Frederiksen *et al.* 2008) and the philopatric behaviour of the surviving adults and chicks (Aebischer 1995; Barlow *et al.* 2013) might be sufficient to retain the colony's genetic diversity.

Geographical distance between shag colonies had a strong influence on the extent of population genetic structure, but only when cross-sea distance was considered. This outcome highlights the role of continental land masses as a physical barrier that promotes isolation between shag populations. Shags are coastal seabirds with limited foraging ranges, usually within a few kilometres during breeding and postbreeding periods, although they can fly a few hundred kilometres between breeding and postbreeding grounds (Sponza *et al.* 2013; Bogdanova *et al.* 2014; Grist *et al.* 2014). Occasional long-distance, cross-sea movements have been also reported for adults and juveniles prior to the colony recruitment (Wanless & Harris 2004). Following crashes, juvenile shags have been shown to move further than during noncrash years (Potts 1969), while adults could emigrate to other colonies (Velando & Freire 2002; Thanou 2013). These events, even if they happen rarely, might promote gene flow among populations, despite high levels of natal and breeding philopatry (Barlow *et al.* 2013). However, long-distance dispersal seems unable to erase the overall genetic structure, and thus, breeding immigrants more often originate from geographically close populations, probably of the same genetic cluster.

Finally, processes such as sex-biased dispersal may affect population genetic structure. A few cases of adjacent shag populations sharing common mtDNA haplotypes despite their significant nuclear divergence (high F_{ST} values) were found within the N. Atlantic and the Aegean, while evidence of a female-biased dispersal to new territories was reported in years with unfavourable weather conditions (Baros *et al.* 2013). Female-biased dispersal within a close range around natal colonies may act to homogenize genetic variation, but further research is needed to clarify whether long-distance female immigration could have a strong effect on the genetic structuring of shag populations.

Taxonomy and conservation

Microsatellite data provided evidence of genetic differentiation between Atlantic and Mediterranean shags, but the transition cline in variation between *P. a. aristotelis* and *P. a. desmarestii* moved to the north and was located among the admixed Spanish populations. Thus, currently described morphological subspecies appear to be paraphyletic and the morphological differences between them might reflect local adaptations. Representation of African and Black Sea shags in future phylogenetic work would help infer the centre of the species' diversification and resolve taxonomic implications.

The North Atlantic and the eastern Mediterranean shags represent two evolutionarily significant units (ESUs), while the Spanish/west Mediterranean populations are probably better described by a meta-population model. Molecular data suggest that past bottlenecks are relatively rare and stable populations have probably persisted throughout the species' range. During recent decades, however, several Atlantic and Mediterranean populations have declined (BirdLife International 2015). Thus, focusing conservation strategies on the specified ESUs could be effective, especially considering that they are distinct units of a monotypic, European endemic genus (Kennedy & Spencer 2014).

Conclusions

The complex phylogeographical history of the shag corroborates the effect of the European palaeogeography on the diversification of marine organisms and particularly the effect of Pleistocene glacial cycles on restricted marine areas, such as the Mediterranean Sea. Moreover, it highlights the importance of lesser studied northern refugia along with well-known southern ones and the role of contemporary ecological barriers within marine areas. As expected, philopatry and local adaptation to specific marine habitat features act to maintain population genetic structure in this seabird. However, neither physical nor nonphysical barriers seem to prevent rare episodes of long-distance dispersal typically associated with unfavourable conditions. This flexibility in dispersal behaviour and tolerance of a wide range of climate conditions (e.g. temperate vs. near-arctic refugia) suggest that different populations may respond individually to habitat and climatic changes, ensuring species survival at least in the case of the shag.

Acknowledgements

We wish to thank everyone who provided samples or facilitated sampling at all colonies, most particularly M. Newell (Isle of May), R. Duncan and C. Jones (Bullers of Buchan), R. Sellers and M. Oksien (Badbea), R. L. Swann (North Sutor and Canna), The National Trust and C. Redfern (Staple Island), O. Merne (Ireland's Eye), S. Newton (Lambay), B. Olsen (Skúvoy), A. Petersen (Flatey), R. Barrett (Hornøya), S. Christensen-

Dalsgaard (Anda), T. Anker-Nilssen (Røst), G. Bangjord (Melstein), S-H. Lorentsen (Sklinna), A. Follestad (Kjøer), P. Yésou, B. Cadiou and J. Nisser (Île de Béniguet), A. Velando (Galicia), J. Hidalgo (Vizcaya), J-C. Thibault (Corsica), A. Christidis, M. Koutrakis and G. Markianos (Xironisi), J. Fric (Sporades and Fournoi), D. Kafentzi (Tilos), P. Lymberakis and S. Xirouchakis (Evvoia and Crete) and B. Cimador and J. Kralj (Oruda and Brijuni). We are also grateful to the Hellenic Ornithological Society (HOS) who supported fieldwork in the Aegean colonies through the LIFE07 NAT/GR/00285 project and the Natural History Museum of Crete for sample loans. We particularly thank Jane Reid and three anonymous reviewers for helpful comments on an earlier version of the manuscript. ET was supported by the Career Development Bursary (2014) of the British Ornithologists' Union (BOU). All samples were collected under appropriate licences in accordance with national legal, ethical and welfare regulations.

References

Aebischer NJ (1986) Retrospective investigation of an ecological disaster in the shag, Phalacrocorax aristotelis – a general method based on long-term marking. *Journal of Animal Ecology*, **55**, 613–629.

Aebischer NJ (1995) Philopatry and colony fidelity of Shags *Phalacrocorax aristotelis* on the east-coast of Britain. *Ibis*, **137**, 11–18.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts, USA.

Avise JC, Nelson WS, Bowen BW, Walker D (2000) Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty tern (*Sterna fuscata*). *Molecular Ecology*, **9**, 1783–1792.

Bandelt H-J, Forster P, Röhl A (1999) Median-Joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barlow EJ, Daunt F, Wanless S, Álvarez D, Reid JM, Cavers S (2011) Weak large-scale population genetic structure in a philopatric seabird, the European Shag *Phalacrocorax aristotelis*. *Ibis*, **153**, 768–778.

Barlow EJ, Daunt F, Wanless S, Reid JM (2013) Estimating dispersal distributions at multiple scales: within-colony and among-colony dispersal rates, distances and directions in European shags *Phalacrocorax aristotelis*. *Ibis*, **155**, 762–778.

Baros Á, Álvarez D, Velando A (2013) Climate influences fledgling sex ratio and sex-specific dispersal in a seabird. *PLoS One*, **8**, e71358.

Berger A (1984) Accuracy and frequency stability of the Earth's orbital elements during the Quaternary. In: *Milankovitch and Climate, Part 1* (eds Berger A, Imbrie J, Hays J, Kukla G, Saltzmann B), pp. 3–39. Reidel, Dordrecht.

BirdLife International (2015) *European Red List of Birds*. Office for Official Publications of the European Communities, Luxembourg.

Bogdanova MI, Wanless S, Harris MP *et al.* (2014) Among-year and within-population variation in foraging distribution of European shags *Phalacrocorax aristotelis* over two decades: implications for marine spatial planning. *Biological Conservation*, **140**, 292–299.

Bonfiglio L, Mangano G, Marra AC, Masini F, Pavia M, Petruso D (2002) Pleistocene Calabrian and Sicilian bioprovinces. *Geobios*, **24**, 29–39.

Cagnon C, Lauga B, Hémery G, Mouchés C (2004) Phylogeographic differentiation of storm petrels (*Hydrobates pelagicus*) based on cytochrome b mitochondrial variation. *Marine Biology*, **145**, 1257–1264.

Calderón L, Quintana F, Cabanne GS, Lougheed SC, Tubaro PL (2014) Phylogeography and genetic structure of two Patagonian shag species (Aves: Phalacrocoracidae). *Molecular Phylogenetics and Evolution*, **72**, 42–53.

Coll M, Piroddi C, Steenbeek J *et al.* (2010) The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. *PLoS One*, **5**, e11842.

Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Duffie CV, Glenn TC, Vargas FH, Parker PG (2009) Genetic structure within and between island populations of the flightless cormorant (*Phalacrocorax harrisi*). *Molecular Ecology*, **18**, 2103–2111.

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Frederiksen M, Daunt F, Harris MP, Wanless S (2008) The demographic impact of extreme events: stochastic weather drives survival and population dynamics in a longlived seabird. *Journal of Animal Ecology*, **77**, 1020–1029.

Friesen VL (2015) Speciation in seabirds: why are there so many species…and why aren't there more? *Journal of Ornithology*, **156**, 27–39.

Friesen VL, Burg TM, McCoy KD (2007) Mechanisms of population differentiation in seabirds. *Molecular Ecology*, **16**, 1765–1785.

Gay L, Neubauer G, Zagalska-Neubauer M *et al.* (2007) Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact. *Molecular Ecology*, **16**, 3215–3227.

Gómez-Diaz E, González-Solis A, Peinado MA, Page RDM (2006) Phylogeography of the *Calonectris* shearwaters using molecular and morphometric data. *Molecular Phylogenetics and Evolution*, **41**, 322–332.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices Version 2.9.3. http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Grist H, Daunt F, Wanless S *et al.* (2014) Site fidelity and individual variation in winter location in partially migratory European Shags. *PLoS One*, **9**, e98562.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Kennedy M, Spencer HG (2014) Classification of the cormorants of the world. *Molecular Phylogenetics and Evolution*, **79**, 249–257.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science*, **277**, 1666–1669.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Krijgsman W, Hilgen FJ, Ray I, Sierro FJ, Wilson DS (1999) Chronology, causes and progression of the Messinian salinity crisis. *Nature*, **400**, 652–655.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Librado P, Rozas J (2009) DnaSP v5 – a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics Applications Note*, **25**, 1451–1452.

Liebers D, de Knijff P, Helbig A (2004) The herring gull complex is not a ring species. *Proceedings of the Royal Society of London Series B Biological Sciences*, **271**, 893–901.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Marion L, Le Gentil J (2006) Ecological segregation and population structuring of the Cormorant *Phalacrocorax carbo* in Europe, in relation to the recent introgression of continental and marine subspecies. *Evolutionary Ecology*, **20**, 193–216.

Mercer DM, Haig SM, Roby DD (2013) Phylogeography and population genetic structure of double-crested cormorants (*Phalacrocorax auritus*). *Conservation Genetics*, **14**, 823–836.

Merji R, Lo Brutto S, Kalthoum O, Hassine B, Arculeo M (2009) A study on *Pomatoschistus tortonesei* Miller 1968 (Perciformes, Gobiidae) reveals the Siculo-Tunisian Strait (STS) as a breakpoint to gene flow in the Mediterranean basin. *Molecular Phylogenetics and Evolution*, **53**, 596–601.

Mlíkovský J (2009) Evolution of the Cenozoic marine avifaunas of Europe. *Annalen des Naturhistorischen Museums in Wien Serie A für Mineralogie und Petrographie, Geologie und Paläontologie, Anthropologie und Prähistorie*, **111**, 357–373.

Morris-Pocock JA, Taylor SA, Brit TP *et al.* (2008) Population genetic structure in Atlantic and Pacific Ocean common murres (*Uria aalge*): natural replicate tests of post-Pleistocene evolution. *Molecular Ecology*, **17**, 4859–4873.

Morris-Pocock JA, Steeves TE, Estela FA, Anderson DJ, Friesen VL (2010) Comparative phylogeography of brown (*Sula leucogaster*) and red-footed boobies (*S. sula*): the influence of physical barriers and habitat preference on gene flow in pelagic seabirds. *Molecular Phylogenetics and Evolution*, **54**, 883–896.

Moum T, Árnason E (2001) Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. *Molecular Ecology*, **10**, 2463–2478.

Newton I (2003) *The Speciation and Biogeography of Birds*. Academic Press, Amsterdam.

Parducci L, Jørgensen T, Tollefsrud MM *et al.* (2012) Glacial survival of boreal trees in northern Scandinavia. *Science*, **335**, 1083–1085.

Peakall R, Smouse PE (2012) GenAlex 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*, **28**, 2537–2539.

Pérez-Losada M, Nolte MJ, Crandall KA, Shaw PW (2007) Testing hypotheses of population structuring in the Northeast Atlantic Ocean and Mediterranean Sea using the common cuttlefish *Sepia officinalis*. *Molecular Ecology*, **16**, 2667–2679.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Potts GR (1969) The influence of eruptive movements, age, population size and other factors on the survival of the shag (*Phalacrocorax aristotelis* (L.)). *Journal of Animal Ecology*, **38**, 53–102.

Potts GR, Coulson JD, Deans I (1980) Population dynamics and breeding success of the shag, *Phalacrocorax aristotelis*, on the Farne Islands, Northumberland. *Journal of Animal Ecology*, **49**, 465–484.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quéré N, Desmarais E, Tsigenopoulos CS, Belkhir K, Bonhomme F, Guinand B (2012) Gene flow at major transitional areas in sea bass (*Dicentrarchus labrax*) and the possible emergence of a hybrid swarm. *Ecology and Evolution*, **2**, 3061–3078.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A (2014) *FigTree. Tree figure drawing tool. Version 1.4.2*. Available from: http://tree.bio.ed.ac.uk/software/figtree.

Rambaut A, Drummond AJ (2009) *Tracer. Ver. 1. 5.[Computer software]*. Edinburgh, UK, Institute of Evolutionary Biology, University of Edinburgh. Available from: http://tree.bio.ed.ac.uk/software/tracer.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rawlence NJ, Till CE, Scofield RP *et al.* (2014) Strong phylogeographic structure in a sedentary seabird, the Stewart Island Shag (*Leucocarbo chalconotus*). *PLoS One*, **9**, e90769.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3 – Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematics Biology*, **61**, 539–542.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Sponza S, Cosolo M, Kralj J (2013) Migration patterns of the Mediterranean Shag *Phalacrocorax aristotelis desmarestii* (Aves:

Pelecaniformes) within the northern Adriatic Sea. *Italian Journal of Zoology*, **80**, 1–12.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
Steeves TE, Anderson DJ, Friesen VL (2005) The Isthmus of Panama: a major physical barrier to gene flow in a highly mobile pantropical seabird. *Journal of Evolutionary Biology*, **18**, 1000–1008.
Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London Series B Biological Sciences*, **277**, 661–671.
Taberlet P, Fumagalli L, Wust-Saucy AG, Cossons JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
Tigano A, Damus M, Birt TP, Morris-Pocock JA, Artukhin YB, Friesen VL (2015) The Arctic: glacial refugium or area of secondary contact? Inference from the population genetic structure of the Thick-Billed Murre (*Uria lomvia*), with implications for management. *Journal of Heredity*, **106**, 238–246.
Thanou E (2013) *Reproductive ecology and genetic structure of the European Shag [Phalacrocorax aristotelis (L., 1761)] in the Aegean, Greece*. PhD Thesis, University of Patras, Patras, Greece.
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.
Velando A, Freire J (2002) Population modelling of European Shags (*Phalacrocorax aristotelis*) at their southern limit: conservation implications. *Biological Conservation*, **107**, 59–69.
Wanless S, Harris MP (2004) European Shag. In: *Seabird Populations of Britain and Ireland* (eds Mitchell PI, Newton SF, Ratcliffe N, Dunn TE). Seabird Populations of Britain and Ireland, London.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Yu Y, Harris AJ, He XJ (2010) S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*, **56**, 848–850.
Yu Y, Harris AJ, Blair C, He XJ (2015) RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**, 46–49.
Zotier R, Bretagnolle V, Thibault JC (1999) Biogeography of the marine birds of a confined sea, the Mediterranean. *Journal of Biogeography*, **26**, 297–313.

E.T., S.C. and F.D. designed the study. E.T., S.S., E.B., S.W. and F.D. contributed to the data set. E.T. undertook all the laboratory and molecular data analyses, with the help of A.P. in microsatellite genotyping and scoring. E.T. and S.C. drafted the manuscript, and all other authors contributed to manuscript revisions. All authors have approved the manuscript for publication.

Data accessibility

MtDNA sequences with GenBank Accession nos: KY354307–KY354334 (CR1) and KY354335–KY354344 (ND2).

Raw microsatellite data in the Supporting Information (Table S6).

Final DNA sequence assembly FASTA files, BEAST and STRUCTURE input files Dryad Digital Repository https://doi.org/10.5061/dryad.b85pj.

Supporting information

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

Fig. S1. Inference of the number of genetic clusters (*K*) estimated for (a) all studied samples and populations, (b) all samples from the Atlantic populations and (c) all samples from the Mediterranean populations.

Appendix S1. Additional information on primer sequences & PCR conditions (Table S1), substitution models checked for the BI analysis (Table S2), substitution rates checked for the molecular clock analysis (Table S3), mtDNA sequences and haplotypes used in the previous analyses (Table S4), estimates of mtDNA genetic distances (Table S5), raw microsatellite genotypes (Table S6), estimates of microsatellite Fst and Rst values (Table S7) and a list of the respective references.