Mitochondrial DNA Sequence Evolution and Phylogeny of the Atlantic Alcidae, Including the Extinct Great Auk (*Pinguinus impennis*)

Truls Moum, Ulfur Arnason,† and Einar Árnason‡*

*Department of Molecular Biotechnology, Institute of Medical Biology, University of Tromsø, Norway; †Department of Genetics, Division of Evolutionary Molecular Systematics, University of Lund, Sweden; and ‡Institute of Biology, University of Iceland, Reykjavik

The Atlantic auk assemblage includes four extant species, razorbill (*Alca torda*), dovekie (*Alle alle*), common murre (*Uria aalge*), and thick-billed murre (*U. lomvia*), and one recently extinct species, the flightless great auk (*Pinguinus impennis*). To determine the phylogenetic relationships among the species, a contiguous 4.2-kb region of the mitochondrial genome from the extant species was amplified using PCR. This region included one ribosomal RNA gene, four transfer RNA genes, two protein-coding genes, the control region, and intergenic spacers. Sets of PCR primers for amplifying the same region from great auk were designed from sequences of the extant species. The authenticity of the great auk sequence was ascertained by alternative amplifications, cloning, and separate analyses in an independent laboratory. Phylogenetic analyses of the entire assemblage, made possible by the great auk sequence, fully resolved the phylogenetic relationships and split it into two primary lineages, *Uria* versus *Alle*, *Alca*, and *Pinguinus*. A sister group relationship was identified between *Alca* and *Pinguinus* to the exclusion of *Alle*. Phylogenetically, the flightless great auk originated late relative to other divergences within the assemblage. This suggests that three highly divergent species in terms of adaptive specializations, *Alca*, *Alle*, and *Pinguinus*, evolved from a single lineage in the Atlantic Ocean, in a process similar to the initial adaptive radiation of alcids in the Pacific Ocean.

Introduction

The family Alcidae is a group of specialized seabird species that capture their prey by wing-propelled diving. Alcids probably originated in the Pacific Ocean (reviewed by Bédard 1985). Today, 22 alcid species inhabit the temperate and cold regions of the Atlantic and Pacific oceans (Gaston and Jones 1998, pp. 3–58). Eighteen species are found in the Pacific Ocean, but only six in the Atlantic Ocean, and two of them, the *Uria* species, are also found in the Pacific. Among the 22 extant species, six major species groups have been identified that are approximately equidistantly related in molecular genetic terms, suggesting a rapid adaptive radiation in early alcid evolutionary history (Moum et al. 1994; Friesen, Baker, and Piatt 1996). Only one of the groups includes exclusive Atlantic genera, *Alca*, *Alle*, and *Pinguinus*. The evolution of this group is consequently thought to have taken place mainly in the Atlantic Ocean (Udvardy 1963; Bédard 1985). The Atlantic lineage consists of four extant species: razorbill (*Alca torda*), dovekie (*Alle alle*), common murre (*Uria aalge*), and thick-billed murre (*U. lomvia*) (Strauch 1985; Moum et al. 1994; Friesen, Baker, and Piatt 1996). In the early 19th century another alcid, the great auk (*Pinguinus impennis*), still inhabited the North Atlantic waters. The great auk represented a rare modern example of a flightless bird in the Northern Hemisphere (Bengtson 1984; Grieve 1885, pp. 1–26). These large, flightless seabirds were hunted down by humans and became extinct in the 1840s. Morphological analyses have suggested that the great auk was closely related to the razorbill (Dawson 1920; Storer 1945; Strauch 1985).

Alcid relationships have previously been assessed using both morphological and molecular analyses, but the great auk has not been previously subjected to molecular studies. Also, the phylogenetic position of the dovekie has remained uncertain (Storer 1945; Strauch 1985; Moum et al. 1994; Friesen, Baker, and Piatt 1996). Here, we address the phylogeny of all the Atlantic alcids on the basis of more extensive sequence data than was available in previous molecular analyses and by including the great auk. Several workers have emphasized the possible limitations in phylogenetic analysis when the size of the molecular data set is small (Faith 1990; Nei, Kumar, and Takahashi 1998). Another problem related to this is that phylogenetic inference is often based on single genes, not taking into account a possible heterogeneity among loci. Such limitations are often encountered in studies of old materials, because of the inherent problems associated with the analysis of partially degraded DNA. The utility of retrieving larger amounts of sequence data from ancient DNA was recently demonstrated in studies of ratite evolution (Cooper et al. 2001; Haddrath and Baker 2001). Thus, the purpose of the present study was to characterize mitochondrial DNA (mtDNA) sequence evolution and resolve the phylogeny of Atlantic auks, including the great auk, on the basis of a comprehensive sampling of regions from the mitochondrial genome with different putative selective constraints.

Key words: mitochondrial DNA, sequence evolution, Atlantic auk assemblage, adaptive radiation, alcid phylogeny, *Pinguinus impennis*.

Address for correspondence and reprints: Truls Moum, Department of Sheep and Goat Research, Norwegian School of Veterinary Science, Kyrkjevegen 332/334, N-4325 Sandnes, Norway. E-mail: truls.moum@veths.no.

Mol. Biol. Evol. 19(9):1434–1439. 2002

Materials and Methods

Tissue Samples and DNA Extractions

Fractions enriched for mtDNA were previously isolated from tissue samples of extant alcid species (Moum et al. 1994), using the method described by Jones et al. (1988). The black guillemot (*Cepphus grylle*), an alcid

found in the Atlantic Ocean but not belonging to the Atlantic species group (Strauch 1985; Moum et al. 1994; Friesen, Baker, and Piatt 1996), was included as an outgroup.

Tissue was obtained from a great auk caught in Iceland in 1821. The specimen was stuffed and is presently in the possession of the Icelandic Natural History Museum in Reykjavik (Petersen 1995). A contour feather from the tail, as well as a small piece of skin from the abdomen, were removed from the specimen. A 3-mm piece of the feather and approximately 4 mg of skin were used for separate DNA extractions at our laboratory in Tromsø, Norway. Later, two skin samples of similar size were removed and analyzed at Lund University, Sweden. This laboratory, specializing in mtDNA sequencing, has not previously been involved in work on alcids or alcid DNA. The extractions and analyses done here served as controls for contamination, in particular as controls for intralaboratory contamination.

Samples were moistened with 15 μl reagent B from the Nucleon HT kit (Amersham Pharmacia Biotech) and finely chopped. Reagent B was then added to bring the volume up to 350 μl. Samples were proteinase K treated for 46 h at 50°C with occasional agitation. Proteinase K was added to a final concentration of approximately 1 μg/μl (18 μl proteinase K [20 mg/ml]). After treatment for 21 h an additional 8 μl proteinase K was added. DNA extraction was completed with the Nucleon HT kit following the manufacturer's instructions, and each of the resulting DNA pellets was dissolved in 60 μl 1 mM Tris–HCl pH 9.0. Extractions were carried out using dedicated equipment and reagents. Appropriate controls were run to monitor contamination (Handt et al. 1994). Specifically, extraction blanks (negative controls) without tissue were treated in parallel through all the steps of DNA isolation, and PCR controls were run in all subsequent amplification reactions.

Polymerase Chain Reactions, Primers, and Sequencing

A contiguous 4.2-kb region of the mitochondrial genome was targeted for sequence analyses, including a partial NADH dehydrogenase subunit 5 gene (*nd5*) and complete cytochrome *b* (*cyt b*), $tRNA^{Thr}$, $tRNA^{Pro}$, *nd6*, $tRNA^{Glu}$, control region (CR), $tRNA^{Phe}$, and *12S* rRNA genes. Several measures were taken to avoid amplification of nuclear copies of mitochondrial genes (numts). Firstly, fractions enriched for mtDNA were used as templates for PCR in the analyses of extant alcid species. Secondly, the capacity of the expand high-fidelity (EHF) PCR system (Boehringer Mannheim) was exploited to amplify long DNA fragments. In this case, PCR conditions were adjusted according to the manufacturer's recommendations to allow amplification of mtDNA fragments up to 4.5 kb. Thirdly, sequence analyses of all species were replicated and the sequences confirmed from alternative amplifications using different primer pairs. A combination of conserved primers and primers designed to match mtDNA sequences of alcids in general or particular species was employed in PCR and sequencing reactions. Sequences were obtained from two or more individuals of each extant species. To minimize the risk of amplifying DNA of human, bacterial, or fungal origin, primers used to amplify great auk DNA were based on sequence information of allied species.

Great auk mtDNA was amplified using AmpliTaq Gold polymerase (Perkin–Elmer) or the EHF system, in conformity with the methods used for the other species. PCR was typically performed in 25-μl reaction volumes containing 2.0–3.0 mM $MgCl_2$, 0.2 mM dNTPs, 1.0 μM of each primer, 0.8 U AmpliTaq Gold or 1.3 U Expand enzyme mix, and 2–4 μl of the great auk DNA solution. Typical PCR conditions using AmpliTaq Gold were 92°C for 10 min, followed by 40 cycles of 92°C, 55–60°C, and 72°C for 30, 30, and 45 s, respectively. Cycling parameters for the EHF reactions were 92°C for 2 min, 10 cycles of 92°C, 55–60°C, and 72°C for 20, 30, and 45 s, respectively, followed by 25 cycles of 92°C, 58°C, and 70°C for 15, 30, and 45 s, respectively, including a 10-s increment per cycle for the elongation step. Excess primers and nucleotides were enzymatically removed from amplification reactions, using a combination of exonuclease I and shrimp alkaline phosphatase (Amersham Pharmacia Biotech). Amplified DNA was then added as a template to cycle sequencing reactions, using the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Pharmacia Biotech). The nucleotide analog dITP was used throughout to avoid compression artifacts.

At the University of Tromsø 13 primer pairs were used to amplify approximately 4.2 kb of the great auk mitochondrial genome in overlapping sections. All amplifications were performed on a PTC-200 thermal cycler (MJ Research). Amplification products from the variable 5′ end of the great auk CR were ligated into the *Sma*I site of a pUC18 vector, using the Sure Clone Ligation kit (Amersham Pharmacia Biotech), and the sequence was determined from three individual clones. The independent replication of great auk DNA analyses at Lund University was performed using 16 freshly synthesized primers. Four conserved primers, i.e., primers that were likely to amplify most avian mtDNAs, were used. The remaining primers were designed to match more variable loci from the putative great auk sequence collected in Tromsø, including four primers at positions that differed from those of the previous analysis. Amplifications were performed with AmpliTaq Gold polymerase on a Robocycler Gradient 96 thermal cycler (Stratagene) and the products sequenced on both strands using the ABI automated sequencing system (Applied Biosystems).

Sequence and Phylogenetic Analyses

Multiple sequence alignments were constructed by the GCG software package (Genetics Computer Group 1999) and manually adjusted by eye. The maximum likelihood methods for combined analyses of multiple genes (Yang 1996) implemented in the PAML package (Yang 1998) were used to estimate critical parameters in the Kimura model of molecular evolution (Kimura 1980). Protein-coding genes were divided into three se-

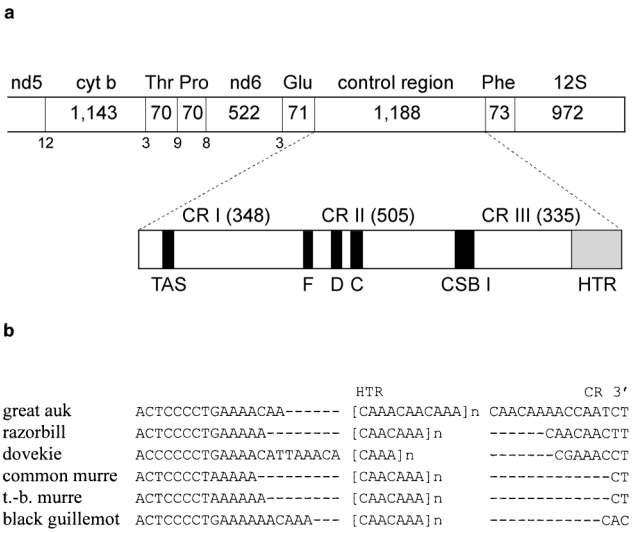

FIG. 1.—Organization and sequence characteristics of great auk mtDNA. *a,* Schematic representation of a 4,258-bp mtDNA region in great auk. The length of each gene is indicated, and lengths of the intergenic spacers are given below the gene junctions. For designation of tRNAs the corresponding three-letter amino acid code is used. A representation of the control region with conserved boxes F, D, C, and conserved sequence block I (CSB I) is given below. The heteroplasmic tandem repeat (HTR) region and a possible TAS are indicated. *b,* Multiple sequence alignment of the 3′ end of the control region showing the position and sequence motifs of HTRs in CR III of six alcid species. Dashes represent gaps. HTR sequence motifs are shown in brackets, and *n* designates the variable number of repeats found within single individuals, a condition known as heteroplasmy. The 3′ end of the CR is indicated.

quence regions consisting of first, second, or third codon positions only (Yang 1998). *Cyt b* and *nd5* were treated jointly, but separately from the light strand–encoded *nd6.* The four tRNA genes were treated jointly. A common three-partite division adopted for the CR was defined as follows: a variable 5′ region (CR I) of 300–400 bp where synthesis of nascent heavy strands pauses at termination-associated sequences (TASs; Doda, Wright, and Clayton 1981); a more conserved middle region (CR II) of approximately 500 bp with several conserved sequence motifs; and a variable-length 3′ region (CR III) which comprises the origin of heavy strand replication, the promoters for transcription, and a conserved sequence block (CSB I) possibly associated with the initiation of heavy strand synthesis (Brown et al. 1986; Clayton 1991; fig. 1*a*). Thus, parameter estimation under the Kimura model was based on 11 sequence regions with putative different primary function and selective constraints, i.e., *12S;* tRNA genes; CR I; CR II; CR III; first, second, and third codon positions of *cyt b–nd5;* and first, second, and third codon positions of *nd6.* The evolutionary rate (c) of *12S* was set at unity and other rates estimated relative to this. The other parameters estimated were the transition-transversion rate ratio (κ) and the shape parameter (α) of the gamma distribution (Yang 1996).

Phylogenetic relationships were analyzed using the maximum likelihood method (Felsenstein 1981) of the PHYLIP package (DNAML; Felsenstein 1993). Sequence data were sampled for phylogenetic analyses according to several alternative criteria to test for consistency. Transition-transversion ratios were adopted from PAML, and taxon input order was varied 10 times using the jumble option in DNAML. Further settings for the analysis presented in figure 2 were determined by maximizing the likelihood value in iterative runs. An ln likelihood of −8,282.86 was obtained using the maximum number of categories allowed in DNAML, i.e., nine categories with the following rates (and probabilities): 0 (0.6), 1 (0.05), 2 (0.05), 3 (0.05), 4 (0.05), 5 (0.05), 10 (0.05), 20 (0.05), and 40 (0.05). The expected length of a patch of sites having the same rate was set at 5.0. The phylogenetic tree was drawn using the TREEVIEW application (Page 1996).

Table 1
Parameters in the Kimura Model

Segment	c	κ	α	Base Pairs
12S rRNA	1	22.96	0.130	962
tRNAs	2.38	60.67	0.113	285
CR I	3.01	4.48	0.934	230
CR II	1.16	4.95	0.216	476
CR III	1.76	8.05	0.794	216
cyt b–nd5				
First codon position	0.59	25.44	0.175	419
Second codon position	0.10	5.53	0.005	419
Third codon position	3.93	10.82	1.557	419
nd6				
First codon position	1.48	33.10	0.199	173
Second codon position	0.27	10.82	0.067	173
Third codon position	4.11	15.29	8.350	173

NOTE.—Estimated parameters in the Kimura (1980) model of molecular evolution for six Atlantic alcids were based on 3,945 homologous positions. Parameter designations: c = relative evolutionary rate; κ = transition-transversion rate ratio; α = shape parameter of the gamma distribution.

Results

Homologous mtDNA sequences of approximately 4.2 kb were compiled for the black guillemot and all species of the Atlantic auk lineage, using novel and previously published sequence data (Moum et al. 1994). Parameters were estimated using the Kimura model of molecular evolution, on the basis of 3,945 homologous nucleotide positions among six species (table 1).

We recovered a contiguous sequence of 4,258 bp from the great auk mitochondrial genome in Tromsø. Partial sequences of the great auk *12S, tRNA^{Phe},* CR, *nd6, tRNA^{Pro}, tRNA^{Thr}, cyt b,* and *nd5,* a total of 2,389 bp, were independently determined at Lund University. Great auk sequences from laboratories in Tromsø and Lund were in complete agreement. Most of the sequences (>95%) were determined on both strands.

Great auk DNA extractions were tested for contaminating DNA by amplifications using conserved rRNA primers. The skin preparations gave avian sequences only, whereas amplification of the feather preparation produced a mixed sequence of human and avian origin. Primers based on sequence information from allied species yielded the putative great auk sequences from both tissue extractions. Fragments of ~450 bp or shorter

were successfully amplified from feather DNA, whereas attempts to amplify fragments of ~530 bp or more were ineffectual. MtDNA fragments ranging in size from approximately 300–700 bp were readily amplified from the skin preparation. The greatly reduced yield of a 798-bp amplification from *cyt b* was still amenable to sequence analysis but indicated that longer amplifications were unlikely to be successful and, thus, were not attempted. It is not obvious how high-fidelity enzymes will affect the copying of damaged bases in old DNA. Thus, AmpliTaq Gold polymerase was used to amplify most of the great auk sequences. However, great auk DNA was also amplified using the EHF enzyme mix for comparison with the sequences obtained by using the conventional polymerase. For this particular specimen at least, identical sequences were retrieved using either PCR system.

Ambiguous sequence readings were obtained from amplifications of great auk CR I. Primer design in domain I of the CR is made difficult by its unusual sequence variability, and there is an extended poly-C tract at the 5′ end, typical of alcid species, that tends to complicate sequence analysis. Notably, Cooper et al. (2001) encountered similar problems in their efforts to sequence the same chromosomal location in the extinct moa *Dinornis giganteus*. In the great auk, amplifications using primers L16.723 (5′-AGT CAT ACC AAT TCC TGC TTG GCT T-3′) and H409 (5′-ATT TCA CGT GAG GAG AAC GAC-3′) revealed two superimposed sequence variants from a specific point of the CR I. However, a strategy involving cloning and sequence-specific primers (cf. Moum and Árnason 2001) unambiguously resolved both sequence variants. A putative mtDNA sequence was identified among the two on the basis of the presence of a conserved TAS element, a typical poly-C tract, and overall sequence homology to other alcids. Moreover, this sequence was supported using an alternative pair of sequence-specific primers at subsequent analyses in Lund. These primers produced the putative mtDNA sequence only (primers H368 [5′-CCT GAA GCT AGT AAC GTA GGA TC-3′] and L16.774 [5′-CAC TGT TGT CAA CTT CAA CTA CAG-3′]). Sequence comparisons of 95 homologous positions among alcid CR Is and the second sequence variant revealed an unequivocal, though distant, relationship, indicating that the sequence could be a nuclear homolog.

Conserved primary sequence characteristics of alcid CRs included sequence blocks F, D, and C in CR II (Southern S., Southern P. J., and Dizon 1988), CSB I in CR III, and a putative TAS in CR I (consensus sequence described by Foran, Hixson, and Brown 1988; fig. 1*a*). A 28-bp insertion was found in CR I of the great auk when compared with other alcids. All Atlantic alcids exhibited the same type of heteroplasmic tandem repeat in the CR as that found in other charadriiform birds (Berg, Moum, and Johansen 1995; Moum and Bakke 2001). The common repeat motif CA_2CA_3 was found in razorbill and murre species, whereas dovekie and great auk harbored shortened and extended motifs, respectively (fig. 1*b*). That heteroplasmy also prevailed in great auk was evident by the range of different-sized fragments produced in amplifications of CR III.

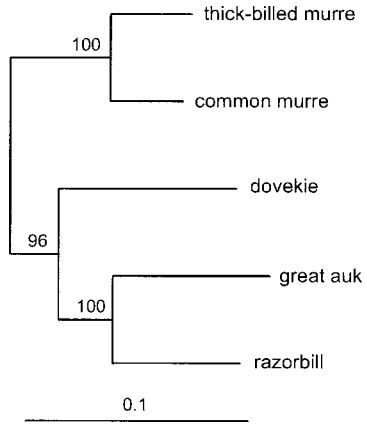

FIG. 2.—Phylogenetic tree of Atlantic alcids derived from maximum likelihood analysis of mtDNA sequences. The analysis was based on 3,140 nucleotide sites showing relatively homogeneous evolutionary rates. The black guillemot was used for outgroup rooting. All branch lengths were significantly positive ($P < 0.01$). Values at the nodes represent bootstrap replication scores based on 100 resamplings.

Sequence comparisons of *nd5* among the alcids showed a thymidine frameshift insertion in the great auk 12 nucleotides from the 3′ end of the gene. The insertion results in replacements of the last three amino acids. Apparently these three residues are evolutionary labile, being poorly conserved in the other alcid species. Also, in all the species the last four nucleotides of the *nd5* are TAGA. The extant species use the AGA stop codon, but after the single basepair insertion upstream the TAG stop codon became available for the great auk, thus not altering the length of the resulting protein.

Phylogenetic analyses supported the placement of dovekie as the sister group of razorbill–great auk. Congeneric murre species were grouped as the other primary lineage among Atlantic alcids. These relationships were supported by analyses of all homologous nucleotide positions ($n = 3{,}945$) and by separate analyses of RNA genes ($n = 1{,}247$), protein-coding genes ($n = 1{,}776$), and the CR ($n = 922$). However, various analyses of the contiguous *nd6*, $tRNA^{Glu}$, and CR I had less support, even when obvious relationships such as that among the congeneric murre species were considered. Base frequencies of this mtDNA region were exceptionally skewed in favor of A and C (37.6% and 34.4%, respectively, in great auk). Consequently, we conjectured that this region might be less useful for phylogenetic inference among alcids, and a phylogenetic analysis based on a more homogeneous data set ($n = 3{,}140$), without *nd6*, $tRNA^{Glu}$, and CR I, was carried out. The tree topology remained unchanged in this analysis, but branch lengths varied less among species, and bootstrap value for the dovekie-razorbill–great auk clade increased from 92 to 96 (fig. 2). Bootstrap values based on the full data of heterogeneous rates were lower than those based on the reduced and more homogeneous data sets. All bootstrap values were significant, and thus the topology was supported by both the full and the reduced data sets.

Discussion

Great Auk DNA Sequence and Patterns of Sequence Evolution

The approximately 180-year-old great auk specimen we analyzed was caught with the intention of preserving it, so it was probably chilled and prepared (salted) relatively quickly, making it a favorable material for present-day studies. Several facts testify to the authenticity of the great auk DNA sequence. PCR product length and amplification efficiency of great auk DNA were inversely related, as expected for amplifications of old, partially degraded DNA samples (Handt et al. 1994). Two samples were analyzed separately, and sequences were confirmed at an independent laboratory on the basis of new DNA extractions from two tissue samples taken from the specimen on a separate occasion. Identical sequences were recovered from amplifications using different primer combinations. Most important, the great auk mtDNA sequence conformed to expectations with regard to sequence relatedness to allied species. Protein-coding gene sequences apparently translate into functional proteins, and RNA genes fold properly into the expected secondary structures, yet the sequences had unique variations not found in the other species. An ambiguous sequence was produced from analysis of the CR I domain in great auk, possibly because of coamplification of the mtDNA and a nuclear homolog. However, a putative mtDNA sequence was determined from cloning of the PCR products, and amplifications using alternative primers confirmed this sequence, thus resolving the ambiguity.

All alcid species, including the great auk, maintain the mitochondrial gene order most commonly observed in birds (Desjardins and Morais 1990; Mindell, Sorenson, and Dimcheff 1998; fig. 1*a*). Although a seemingly obligate existence of intraindividual size variation in CR III makes it impossible to assign a specific length to any alcid CR, the minimum number of repeats in particular individuals is suggested from sequencing gels. In the great auk specimen we analyzed, a minimum of nine repeats seems to be present, which implies a minimum length of 1,188 bp for the entire CR.

As only four extant species are found within the Atlantic auk assemblage, analysis of mtDNA sequence evolution is considerably improved by the great auk sequence data. Estimated parameters in the Kimura model of molecular evolution showed that the evolutionary rates of the first and second codon positions in *nd6* were about 2–3 times those of *cyt b–nd5,* and the rate of the third codon positions exceeded that of the CR sequences (table 1). However, evolutionary rates of CR sequences are underestimated by this method, because of insertions or deletions and the concomitant exclusion of many nucleotides from sequence comparisons. The extent of rate variation among sites varied widely with gene segment. As expected, because of functional constraints the shape of the gamma distribution was highly skewed for RNA genes and the first and second codon positions of protein-coding genes. The alpha parameter approached 1.0 in domain I of the CR and exceeded 1.0 in third codon positions, indicating even rates for most sites in these segments.

Phylogenetic Analyses and Evolutionary Inferences

The affiliation of great auk to razorbill suggested from morphological analyses was corroborated by the molecular data. Furthermore, phylogenetic analyses supported the placement of dovekie as the sister group of razorbill–great auk. These relationships were strongly supported by analyses of the complete data and by most of the analyses performed on each type of mitochondrial genetic element, thus taking into account the apparent evolutionary rate heterogeneity. However, we suspect the contiguous *nd6*, $tRNA^{Glu}$, and CR I to be less appropriate for phylogenetic analyses in alcids. There is a pronounced bias in base composition, and this part of the mitochondrial genome seems to be highly susceptible to mutation, as is evident from the elevated evolutionary rates, the frequent occurrence of insertions or deletions (making alignments unreliable), and the hypervariable primary sequence of CR I in birds, even at the intraspecific (e.g., Wenink, Baker, and Tilanus 1993; Kidd and Friesen 1998; unpublished data) and intraindividual levels (Moum and Bakke 2001).

The current study identified a close relationship between the great auk and the dovekie. This finding was somewhat unexpected considering the fact that these two species are two of the most divergent alcid taxa in terms of adaptive specializations and body size. Bédard (1985) noted that apart from the smaller number of Atlantic species, the sets of alcid species found in the Pacific and Atlantic oceans are remarkably convergent. An initial adaptive radiation of alcids in the Pacific Ocean led to the evolution of a diverse community with small plankton-feeding types (*Aethia*, *Ptychoramphus*), large flightless fish-feeders (the extinct *Mancallas*), and several intermediate forms (e.g., *Fratercula*, *Cepphus*, *Uria*). Evidently, a secondary radiation has taken place within a single lineage in the Atlantic Ocean to produce adaptively equivalent forms: a small plankton-feeding type (dovekie) and a large flightless fish-feeder (great auk), as well as a medium-sized generalist (razorbill). Differentiation in the Pacific probably involved ecological specializations and geographical isolation in refugia (Udvardy 1963). Apparently, speciations that parallel those in the Pacific Ocean also found place in the Atlantic Ocean in spite of the latter's more restrained possibilities for geographical isolation.

Although the literature dealing with the great auk is impressive, no naturalist ever studied the bird in the field, and many aspects of its biology shall remain hazy (Bengtson 1984). Further study of the great auk will be confined to fossils and museum collections. Remains of at least 80 specimens have been preserved in various ways (stuffed, dried skins, ethanol preserved) that make them potentially amenable to DNA sequence analysis. This study represents the first step toward applying population genetic analysis of the great auk to advance our understanding of its ecology and the ultimate reasons for its disappearance.

Data Deposition

The sequences reported in this paper have been deposited in the EMBL database, accession nos. AJ242683–AJ242688.

Acknowledgments

We thank the Natural History Museum of Iceland for access to tissue samples and Ævar Petersen and Thorvaldur Thor Björnsson for help in obtaining the samples. We also thank Anette Gullberg and Peik Haugen for technical assistance and Steinar Johansen, Jaime Garcia-Moreno, and an anonymous reviewer for helpful comments. The Norwegian Research Council supported the study financially.

LITERATURE CITED

BÉDARD, J. 1985. Evolution and characteristics of the Atlantic Alcidae. Pp. 6–19 *in* D. N. NETTLESHIP and T. R. BIRKHEAD, eds. The Atlantic Alcidae. Academic Press, London.

BENGTSON, S. A. 1984. Breeding ecology and extinction of the great auk (*Pinguinus impennis*): anecdotal evidence and conjectures. Auk **101**:1–12.

BERG, T., T. MOUM, and S. JOHANSEN. 1995. Variable numbers of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. Curr. Genet. **27**:257–262.

BROWN, G. G., G. GADALETA, G. PEPE, C. SACCONE, and E. SBISÀ. 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. **192**:503–511.

CLAYTON, D. A. 1991. Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cell. Biol. **7**:453–478.

COOPER, A. C., C. LALUEZA-FOX, S. ANDERSON, A. RAMBAUT, J. AUSTIN, and R. WARD. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature **409**:704–707.

DAWSON, W. L. 1920. An oological revision of the Alciformes. J. Mus. Comp. Zool. **1**:7–14.

DESJARDINS, P., and R. MORAIS. 1990. Sequence and gene organization of the chicken mitochondrial genome. J. Mol. Biol. **212**:599–634.

DODA, J. N., C. T. WRIGHT, and D. A. CLAYTON. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78**:6116–6120.

FAITH, D. P. 1990. Chance marsupial relationships. Nature **345**: 393–394.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1993. PHYLIP: phylogeny inference package. Version 3.5c. University of Washington, Seattle.

FORAN, D. R., J. E. HIXSON, and W. E. BROWN. 1988. Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop synthesis. Nucleic Acids Res. **16**:5841–5861.

FRIESEN, V. L., A. J. BAKER, and J. F. PIATT. 1996. Phylogenetic relationships within the Alcidae (Charadriiformes: Aves) inferred from total molecular evidence. Mol. Biol. Evol. **13**:359–367.

GASTON, A. J., and I. L. JONES. 1998. The auks. Oxford University Press, Oxford.

GENETICS COMPUTER GROUP (GCG). 1999. Wisconsin package. Version 10.0. Genetics Computer Group, Madison, Wis.

GRIEVE, S. 1885. The great auk, or garefowl (*Alca impennis* Linn.). Its history, archeology, and remains. Grange Publishing Works, Edinburgh.

HADDRATH, O., and A. J. BAKER. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proc. R. Soc. Lond. B **268**:939–945.

HANDT, O., M. HÖSS, M. KRINGS, and S. PÄÄBO. 1994. Ancient DNA: methodological challenges. Experientia **50**: 524–529.

JONES, C. S., H. TEGELSTRÖM, D. S. LATCHMAN, and R. J. BERRY. 1988. An improved rapid method for mitochondrial DNA isolation suitable for use in the study of closely related populations. Biochem. Genet. **26**:83–88.

KIDD, M. G., and V. L. FRIESEN. 1998. Sequence variation in the guillemot (Alcidae: *Cepphus*) mitochondrial control region and its nuclear homolog. Mol. Biol. Evol. **15**:61–70.

KIMURA, M. A. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

MINDELL, D. P., M. D. SORENSON, and D. E. DIMCHEFF. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA **95**:10693–10697.

MOUM, T., and E. ÁRNASON. 2001. Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. Mol. Ecol. **10**: 2463–2478.

MOUM, T., and I. BAKKE. 2001. Mitochondrial control region structure and single site heteroplasmy in the razorbill (*Alca torda*; Aves). Curr. Genet. **39**:198–203.

MOUM, T., S. JOHANSEN, K. E. ERIKSTAD, and J. F PIATT. 1994. Phylogeny and evolution of the auks (subfamily *Alcinae*) based on mitochondrial DNA sequences. Proc. Natl. Acad. Sci. USA **91**:7912–7916.

NEI, M., S. KUMAR, and K. TAKAHASHI. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. Proc. Natl. Acad. Sci. USA **95**:12390–12397.

PAGE, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**:357–358.

PETERSEN, Æ. 1995. Brot úr sögu Geirfuglsins. Náttúrufrædingurinn **65**:53–66 [in Icelandic, with English summary].

SOUTHERN, S., P. J. SOUTHERN, and A. E. DIZON. 1988. Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. **28**:32–42.

STORER, R. W. 1945. Structural modifications in the hind limb in the Alcidae. Ibis **87**:433–456.

STRAUCH, J. G. JR. 1985. The phylogeny of the Alcidae. Auk **102**:520–539.

UDVARDY, M. D. F. 1963. Zoogeographical study of the Pacific Alcidae. Pp. 453–470 *in* J. L. GRESSIT, ed. Pacific basin biogeography. Bishop Museum Press, Honolulu.

WENINK, P. W., A. J. BAKER, and M. G. J. TILANUS. 1993. Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpina*). Proc. Natl. Acad. Sci. USA **90**: 94–98.

YANG, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. **42**:587–596.

———. 1998. PAML: phylogenetic analysis by maximum likelihood. Version 1.4. University College, London.

AXEL MEYER, reviewing editor

Accepted April 15, 2002