Sequence Variation in the Guillemot (Alcidae: Cepphus) Mitochondrial Control Region and its Nuclear Homolog

Monica G. Kidd and Vicki L. Friesen

Department of Biology, Queen’s University, Kingston, Ontario, Canada

We describe sequence variation in the mitochondrial control region and its nuclear homolog in three species and seven subspecies of guillemots (Cepphus spp.). Nuclear homologs of the 5’ end of the control region were found in all individuals. Nuclear sequences were ~50% divergent from their mitochondrial counterparts and formed a distinct phylogenetic clade; the mitochondrial-nuclear introgression event must have predated the radiation of Cepphus. As in other vertebrates, the guillemot control region has a relatively conserved central block flanked by hypervariable 5′ and 3′ ends. Mean pairwise interspecific divergence values among control regions were lower than those in other birds. All individuals were heteroplasmic for the number of simple tandem nucleotide repeats (A(n)C) at the 3′ end of the control region. Phylogenetic analyses suggest that black guillemots are basal to pigeon and spectacled guillemots, but evolutionary relationships among subspecies remain unresolved, possibly due to incomplete lineage sorting. Describing molecular variation in nuclear homologs of mitochondrial genes is of general interest in phylogenetics because, if undetected, the homologs may confound interpretations of mitochondrial phylogenies.

Introduction

Characterizing geographic patterns of genetic variation within and among populations is a necessary precursor to understanding mechanisms of population differentiation and speciation. The mitochondrial control region is useful for describing such patterns since it evolves rapidly and because mitochondrial DNA is maternally inherited, does not recombine, and houses selectively neutral variation (Brown, George, and Wilson 1979; Wilson et al. 1985; Moritz, Dowling, and Brown 1987; Gyldensten et al. 1991; Moritz 1994; but see Ballard and Kreitman 1995). Thus, control region sequence variation provides a record of maternal lineages that allows inferences to be made about evolutionary processes such as demographic changes and dispersal. Very few studies describe nucleotide variation in the control region at the intraspecific level in birds (Quinn 1992; Edwards 1993; Wenink, Baker, and Tilanus 1994; Marshall and Baker 1997), and our objective was to gather this information from a morphologically polytypic taxon.

Guillemots (Cepphus spp.) are a genus of pursuit diving seabirds (Charadriiformes: Alcidae). Paleoecological and paleontological information suggests that guillemots first evolved in the Pacific Ocean, then dispersed through the Bering Strait and across the North American Arctic coast into the Atlantic Ocean during the Late Tertiary (Storer 1952; Udvardy 1963). Subsequent isolation of the Atlantic Ocean from the Pacific Ocean by the Bering Land Bridge probably facilitated allopatric divergence of the Atlantic lineage, leading to black guillemots (C. grylle; fig. 1). Emergence of a land bridge from Kamchatka to Hokkaido then may have caused divergence of two Pacific lineages, leading to spectacled guillemots (C. carbo) in the Sea of Okhotsk and pigeon guillemots (C. columba) in the Pacific Ocean. Storer (1952) and Udvardy (1963) suggested that subspecific morphological variation arose much later, probably as a result of isolation of populations during glacial periods of the late Quaternary. Here, we describe sequence variation in the control region and a nuclear homolog of the 5’ end of the control region in 16 birds representing three species and seven subspecies of guillemots. We evaluate the utility of these loci for phylogenetic analysis and expand on what is known of nuclear copies of mitochondrial genes in vertebrates.

Materials and Methods

Samples

Feather, blood, or muscle samples were collected from two adults or chicks in each sampling location (table 1) and were either frozen at −80°C or stored in 95% ethanol at 4°C. Cells were digested with proteinase K in a buffer of 100 mM Tris-Cl (pH 8.0), 10 mM EDTA (pH 8.0), 100 mM NaCl, and 1% sodium dodecyl sulfate overnight at 65°C (Kocher et al. 1989). Proteins and lipids were extracted with phenol saturated with Tris-Cl (pH 8.0) and 24:1 chloroform : isoamyl alcohol. In most cases, DNA was precipitated with 0.2 M sodium acetate and two volumes of cold 95% ethanol, then resuspended in water.

Amplification and Sequencing

We used the polymerase chain reaction (PCR) with versatile and guillemot-specific primers (table 2) to amplify products from total DNA under the following conditions: 10 mM Tris-Cl (pH 8.8), 50 mM KCl, 2 mM MgCl₂, 1 μM each of the light- and heavy-strand primers, 1 mM of an equimolar solution of dNTPs, and 0.5 U Boehringer-Mannheim Taq DNA polymerase. An initial 90-s denaturation at 94°C was followed by 30 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 30 s. PCR products were isolated by electrophoresis through a 1% agarose gel and visualized with ethidium bromide fluorescence under ultraviolet light. PCR bands were excised, and the amplified DNA was purified using Gene Clean II kits (Bio 101, Inc.) according to the manufac-
Fig. 1.—Breeding distribution of the genus Cepphus. (1) C. columba eureka, (2) C. c. adianta, (3) C. c. kaiurka, (4) C. c. columba, (5) C. c. snowi, (6) C. carbo, (7) C. grylle mandtii, (8) C. g. grylle, (9) C. g. islandicus, (10) C. g. faeoreensis, (11) C. g. arcticus, (12) C. g. ultimus. Dashed lines delimit the ranges of species. Adapted from Storer (1952).

Table 1

Numbers and Sampling Locations of Specimens Used in this Study

<table>
<thead>
<tr>
<th>Subspecies</th>
<th>Specimen No.</th>
<th>Tissue Type</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectacled guillemot</td>
<td>CCO-MA-4372</td>
<td>Muscle</td>
<td>Magadan, Russia</td>
</tr>
<tr>
<td></td>
<td>CCO-MA-6055</td>
<td>Muscle</td>
<td></td>
</tr>
<tr>
<td>Pigeon Guillemot</td>
<td>CCA-AL-0054</td>
<td>Muscle</td>
<td>Aleutian Islands, Alaska</td>
</tr>
<tr>
<td></td>
<td>CCA-AL-0055</td>
<td>Muscle</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCE-OR-0001</td>
<td>Feather</td>
<td>Coos Bay, Oregon</td>
</tr>
<tr>
<td></td>
<td>CCE-OR-0002</td>
<td>Feather</td>
<td></td>
</tr>
<tr>
<td>Black guillemot</td>
<td>CGA-SD-0001</td>
<td>Feather</td>
<td>Soderskär, Finland</td>
</tr>
<tr>
<td></td>
<td>CGA-SD-0002</td>
<td>Feather</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGI-FF-002A</td>
<td>Blood</td>
<td>Flatey Island, Iceland</td>
</tr>
<tr>
<td></td>
<td>CGI-FF-002B</td>
<td>Blood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGG-SF-0001</td>
<td>Feather</td>
<td>Sud Fugløya, Norway</td>
</tr>
<tr>
<td></td>
<td>CGG-SF-0002</td>
<td>Feather</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGM-SV-0001</td>
<td>Blood</td>
<td>Svalbard, Norway</td>
</tr>
<tr>
<td></td>
<td>CGM-SV-0002</td>
<td>Blood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGU-DS-2189</td>
<td>Muscle</td>
<td>Nuvuk Island, N.W.T., Canada</td>
</tr>
<tr>
<td></td>
<td>CGU-DS-2190</td>
<td>Muscle</td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Sequences of Primers Used in this Study

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nuclear primers</strong></td>
<td></td>
</tr>
<tr>
<td>ND6</td>
<td>CCTAGAAAAACACCAAATAGGATCAT</td>
</tr>
<tr>
<td>CGL-164</td>
<td>GAGCTAGGACACTGGAATG</td>
</tr>
<tr>
<td>SGH-16500</td>
<td>GCCTGACCGAGGAACCAGA</td>
</tr>
<tr>
<td><strong>Mitochondrial primers</strong></td>
<td></td>
</tr>
<tr>
<td>CGL-001</td>
<td>GGCCTTTTCTCCAAGACCCGCGCTGA</td>
</tr>
<tr>
<td>CGL-56</td>
<td>GIT/CCTCAATACCAATTAATTCAGCA</td>
</tr>
<tr>
<td>CGH-248</td>
<td>ATGGTATTAGTACG (A/T) ATATTGTAATG</td>
</tr>
<tr>
<td>CGH-549</td>
<td>GTATCGGTGAAGTACAAGTTGAGAGG</td>
</tr>
<tr>
<td>CGL-486</td>
<td>AGCCCAACTTGCTCTTTTGCAC</td>
</tr>
<tr>
<td>CGH-1006</td>
<td>TTAATGAAACTCACTGCCGTTTGTAG</td>
</tr>
<tr>
<td>CGH-248</td>
<td>ATGGTATTAGTACG (A/T) ATATTGTAATG</td>
</tr>
<tr>
<td>CGH-549</td>
<td>GTATCGGTGAAGTACAAGTTGAGAGG</td>
</tr>
<tr>
<td>CGL-1256</td>
<td>AGCCGTCTTGGCATCTTCAGTGC</td>
</tr>
</tbody>
</table>

**Note.**—“H” and “L” refer to heavy and light strands, respectively, and numbers refer to the site of the 3′ end of the primer in the guillemot sequence. The prefix “CG” identifies guillemot-specific primers.

* T. P. Birt, unpublished data.


or natural heteroplasmy in repeat copy number, we increased the stringency of the amplification reaction in several ways to rule out slippage mistakes by the Taq enzyme: we increased the annealing temperature to 68°C, decreased the extension time to 15 s, decreased the number of PCR cycles to 25, substituted Taq polymerase with a polymerase with 5–15 times higher fidelity (Vent®; Perler, Comb, and Jack 1992), and adjusted the Mg²⁺ concentration in the reaction buffer.

Since nuclear copies of mitochondrial genes have been reported in other birds (Quinn 1992; Arctander 1995; Friesen and Anderson 1997) and in other vertebrates (Zullo et al. 1991; Collura and Stewart 1995; Zischler et al. 1995; Zhang and Hewitt 1996), we wanted to determine whether amplification products were mitochondrial. We isolated enriched (but not 100% pure) mitochondrial and nuclear DNA fractions from muscle...
samples of four pigeon guillemots using cesium chloride (CsCl) ultracentrifugation (Lansman et al. 1981), then ran a series of test amplifications, ND6 and SGH-16500, CGL-486 and CGL-1006, and versatile primers for one known nuclear locus (MHC class II, unpubl.) and two mitochondrial loci (cytochrome b Kocher et al. [1989] and ATPase 6/8 [H. Walsh, unpublished data]) were used to attempt amplifications of 1:1 through 1:100,000 serial dilutions of the enriched mitochondrial and nuclear DNA fractions. We expected that, with progressive dilution, enriched nuclear fractions should yield product only from nuclear primers, and enriched mitochondrial fractions should yield product only from mitochondrial primers.

Sequence Alignment, Estimation of Divergence, and Phylogenetic Reconstruction

Sequences of nuclear and mitochondrial DNA were aligned using GeneWorks® version 2.4.5 (IntelliGenetics, Inc.) using the default settings; all other sequences were aligned manually using the Eyeball Sequence Editor (Cabot and Beckenbach 1989). Numbers of transitions and transversions and percent sequence divergence were calculated using MEGA version 1.0 (Kumar, Tamura, and Nei 1993). To determine whether divergence between species was significantly different among parts of the control region, one-way analyses of variance were conducted on percent sequence divergence using JMP version 3.0.2 (SAS Institute, Inc. 1994).

In molecular phylogenetic analysis, the region under study must have low levels of homoplasy variation (“multiple hits”), indicated in part by a low proportion of transversions; to determine whether mitochondrial and nuclear sequences had reached phylogenetic saturation, we calculated the observed mean ratio of transitions (s) to transversions (v) and tested whether they differed from ratios expected at phylogenetic saturation. The transition-to-transversion ratio expected at saturation may be estimated from empirical base frequencies (Holmquist 1983):

\[ R_{sv} = \frac{(p_A p_G + p_C p_T)/(p_A + p_G)(p_C + p_T)}{R_{i}} \]

where \( p_i \) is the frequency of base i.

Since the accuracy of different phylogenetic methods is controversial (e.g., Swoford and Olsen 1990; Hillis, Huelsenbeck, and Cunningham 1994), we chose to employ a suite of reconstruction techniques. Maximum-parsimony analyses were performed using midpoint rooting and the branch-and-bound algorithm in PAUP version 3.1.1 (Swoford 1993), with and without an arbitrary 4:1 transversion : transition weighting scheme; bootstrap resampling was used to determine the support for each node. Neighbor-joining (NJ) analyses and standard errors tests using Kimura’s two-parameter genetic distance were conducted with MEGA version 1.0 (Kumar, Tamura, and Nei 1993). Maximum-likelihood (ML) analysis was performed using the DNAML program in PHYLIP version 3.2 (Felsenstein 1989) with the default settings.

Results

Isolation and Identification of the Control Region and its Nuclear Homolog

Our suspicions that we had amplified nuclear DNA were related to the nature of sequence variation observed when amplifying and sequencing using the versatile ND6 and SGH-16500 primers (all sequences have been deposited in GenBank, accession numbers AF027234–AF027249 for nuclear sequences). First, partial sequence from the putative rRNA for glutamic acid (which occurs between the gene for ND6 and the control region in birds; fig. 2A) did not have sufficient internal complementarity to form the correct cloverleaf secondary structure (Desjardins and Morais 1990). Second, sequence from the ND6 and SGH-16500 primers had an order of magnitude less nucleotide diversity than expected if evolutionary constraints on mitochondrial DNA are relaxed, given an approximate time frame of divergence for contemporary guillemots of 2.8 Myr (Storer 1952; unpublished data). Mean pairwise intraspecific sequence divergence (\( p \)) was 0.09% (SD = 0.2%) for the nuclear homolog in black guillemots and zero in both pigeon and spectacled guillemots. All pigeon and spectacled guillemots shared the same genotype, which differed from genotypes of black guillemots by six transitions and three transversions (\( p = 1.9\%, SD = 0.2\% \)). If control region sequences diverge at a rate of 20%/Myr (assuming that the average substitution rate for mtDNA in birds is approximately 2%/Myr and that the substitution rate for the control region is ~10 times the average; Wilson et al. 1985; Quinn 1992; Wenink, Baker and Tilanus 1994), then we would have expected interspecific p to be an order of magnitude higher.

Amplifications of the CsCl-prepared samples confirmed our suspicions. ND6 and SGH-16500 did not amplify products from the four diluted mtDNA fractions, even though versatile primers for the mitochondrial cytochrome b and ATPase 6/8 genes yielded products; however, the ND6 and SGH-16500 primers did yield products from the diluted nuclear fraction, even though the primers for cytochrome b and ATPase 6/8 did not. In contrast, primers CGL-486 and CGL-1006 gave results identical to those of the known mitochondrial primers, yielding product from the diluted mitochondrial fractions but not from the diluted nuclear fractions. We confirmed that ND6 and SGH-16500 amplified nuclear DNA by sequencing the amplification product from the diluted nuclear fraction (data not shown): this template contained virtually no mitochondrial DNA yet produced an amplification product with the same sequence as that from total genomic DNA; therefore, the ND6/SGH-16500 amplification product is most likely of nuclear origin.

Several lines of evidence indicate that the product amplified by primers ND6/SGH-16500 is a homolog of the mitochondrial control region rather than a spurious product. First, sequence homology between the template and the primers (which were designed to anneal to the ND6 gene and a conserved sequence block within the avian mitochondrial control region) was sufficiently
Fig. 2.—Example alignments of sequences of mitochondrial genes and their nuclear homologs in the guillemot. A. The chicken tRNA\textsuperscript{Glu} and the guillemot nuclear homolog. B. The guillemot control region and its nuclear homolog. Identical sites are highlighted. Positions of internal primers are indicated by arrows. The anticodon site in the tRNA and conserved sequence blocks in the control region are underlined. "±" is an indel, and "N" is an unscorable base. Parentheses indicate the AC repeat motif.

In each species, base frequencies in the nuclear homolog and the control region departed significantly from equality (data not shown). The proportions of each base were similar between the nuclear homolog and the control region for each species; as in other mitochondrial genes in other birds, the G content of the L strand was low (e.g., Kocher et al. 1989; Birt et al. 1992; Marshall and Baker 1997).

Patterns of Variation in Control Region Sequences

We obtained sequence of the complete mitochondrial control region for 3 individuals representing each of the three guillemot species and partial sequence (>80% of the total length) for 13 additional individuals representing three species and seven subspecies (GenBank accession numbers for control region sequences are AF027220–AF027233, AF027250–AF027251). The guillemot control region (fig. 2B) is approximately 1,080 bp in length and has conserved sequences characteristic of control regions in other vertebrates. For example, the D box (Southern, Southern, and Dizon 1988) spans positions 486–510 and differs from the chicken sequence by transversions at positions 506–509.
and 507. Conserved sequence block 1 spans positions 831–856 and differs from the chicken sequence (Desjardins and Morais 1990) by a transversion at position 834. Since we did not get complete sequence for part I for all birds and do not consider it appropriate to include the A(n)C repeat in calculations of interspecific substitutional divergence (see below), we focused on 183 bp at the 3’ end of part I, all of part II (549 bp), and the first 155 bp of part III for further comparisons (table 3). For each of the three interspecific comparisons, percent divergence was significantly different among parts I, II, and III (spectacled × pigeon guillemots, $F = 204.8, df = 2,21, P < 0.0001$; spectacled × black guillemots, $F = 28.4, df = 2,57, P < 0.0001$; pigeon × black guillemots, $F = 8587.0, df = 2,117, P < 0.0001$). As in other vertebrates, the guillemot control region has a relatively conserved central region flanked by hypervariable 5′ and 3′ ends (fig. 3). Overall, pigeon and spectacled guillemots share the greatest sequence similarity, and black guillemot sequences are more similar to spectacled guillemot sequences than to pigeon guillemot sequences.

**Phylogenetic Analyses**

Maximum-parsimony analysis of mitochondrial and nuclear sequences produced two monophyletic clades (fig. 4). The mitochondrial sequences provided excellent resolution among species: sequences of spectacled guillemots and pigeon guillemots were more closely related to each other than to those of black guillemots. Phylogenetic relationships among subspecies were not concordant with Storer’s (1952) designation of subspecies according to morphology (with the exception of C. g. islandicus mitochondrial sequences). The nuclear sequences of the black guillemots were monophyletic; however, extensive paraphyly occurred among nuclear sequences of pigeon and spectacled guillemots. The empirical $s/v$ ratio (1.11) of the nuclear copy is significantly greater than the expected ratio at saturation (0.49; $t = 6.317, P < 0.0001$), indicating that the nuclear copy has not reached phylogenetic saturation; however, because it has little variation, it has limited applications for studying genetic differentiation within

---

**Table 3**

<table>
<thead>
<tr>
<th>Part I (183 bp)</th>
<th>Part II (549 bp)</th>
<th>Part III (155 bp)</th>
<th>Total (887 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>$v$</td>
<td>$p$</td>
<td>$s$</td>
</tr>
<tr>
<td>Spectacled × pigeon</td>
<td>6.75</td>
<td>3.00</td>
<td>0.062</td>
</tr>
<tr>
<td>Spectacled × black</td>
<td>(0.89)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Pigeon × black</td>
<td>12.70</td>
<td>8.20</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td>(0.62)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.61)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

*Note.—Roman numerals refer to regions defined by Wenink, Baker, and Tilanus (1994); $s =$ transitions, $v =$ transversions, $p =$ percent divergence.*

---

**FIG. 3.—**Numbers of interspecific substitutions within 25-bp segments of the guillemot control regions.
and among populations of guillemots. Therefore, the remaining analyses were performed on mitochondrial sequences only.

The expected $s/v$ ratio for the control region at saturation is 0.54. The observed $s/v$ ratio exceeded 0.54 in all parts of the control region (one-tailed $t$-test that $s/v = 0.54$, part I: $s/v = 1.81$, $t = 40.992$, $P < 0.0001$; part II: $s/v = 2.37$, $t = 12.047$, $P < 0.0001$; part III: $s/v = 2.00$, $t = 2.717$, $P < 0.004$; overall: $s/v = 2.69$, $t = 39.79$, $P < 0.0001$), indicating that all parts of the control region are appropriate for phylogenetic analysis in guillemots.

There were 77 variable sites in 941 bp of control region sequence, including the 3′ repeat region with repeating units counted as single characters. Of these, 69 were shared by more than one individual and therefore were potentially phylogenetically informative. Since the most plausible mechanism for the evolution of tandemly repeated DNA units is modeled on slippage during replication (Wenink, Baker, and Tilanus 1994) or misalignments of the mitochondrial D-loop strand with multiple termination-associated sequences during replication (Buroker et al. 1990) rather than multiple single-base insertion events, we treated each repeat unit as a single character with binary (presence/absence) or multiple character states. All phylogenetic analyses were performed including and excluding this region; no differences in tree topology resulted, so the repeat region was included in all subsequent manipulations. Further, weighting transversions four times transitions had no effect on topology, so no weighting scheme was used.

The NJ and ML trees for the mitochondrial sequences are shown in figure 5; the parsimony reconstruction provided very little resolution among taxa, and all nodes were also represented in distance-based reconstruction methods. The NJ tree indicated excellent bootstrap and standard errors test support for monophyly of sequences of spectacled and pigeon guillemots; bootstrap values are not true probabilities, but high (≥90) values may be considered strong support for a node (Hillis, Huelsenbeck, and Cunningham 1994; Friesen, Baker, and Piatt 1996). As in figure 4, relationships among control region sequences of the subspecies generally did not reflect taxonomic boundaries based on morphological variation. For example, sequence of one C. g. ultimus from the the Northwest Territories (Canada) grouped with sequence of one C. g. arcticus from Finland with moderate support from bootstrap and standard errors tests. Sequences of black guillemots from Iceland, Finland, and the Northwest Territories formed
Fig. 5.—Phylogenies of control region sequences. See table 1 for sampling locations. a, Neighbor-joining phylogeny. Branch lengths are proportional to Kimura’s two-parameter distance. Numbers above branches indicate bootstrap support; numbers below branches indicate values of support from standard errors tests. b, Maximum-likelihood phylogeny. Branches in bold are significantly longer than 0 at $P < 0.01$.

Discussion
Evolution of the Nuclear Homolog

We positively identified a nuclear homolog of the mitochondrial control region of the guillemot. Although the mechanisms are controversial, nuclear-organellar genomic introgression has been identified in birds (e.g., Quinn 1992; Arctander 1995; Friesen and Anderson 1997) and may be common among other vertebrates (Zullo et al. 1991; Collura and Stewart 1995; Zischler et al. 1995; references in Hu and Thilly 1995; Blanchard and Schmidt 1996; Zhang and Hewitt 1996). This is one of the first surveys of intraspecific sequence variation of a nuclear homolog of a mitochondrial locus (e.g., Edwards 1993; Norman, Moritz, and Limpus 1994; Stewart and Baker 1994; Marshall and Baker 1997). Since all guillemot species possess the nuclear copy, it likely evolved in a lineage ancestral to contemporary Cepphus. Although too many assumptions are required to date the event, the introgression event was likely ancient, as nuclear copies were ~50% divergent from the homologous area of the control region; this value is greater than that of any other reported mitochondrial-nuclear divergence among animals (Zhang and Hewitt 1996). Also, both the monophyletic relationships among the nuclear sequences and the lack of variation suggest that the nuclear copies originated from a single insertion event (Zullo et al. 1991). The nuclear copy has not reached phylogenetic saturation, and as a “molecular fossil” (Collura and Stewart 1995; Zischler et al. 1995) of the control region, it could provide a comparative data set to corroborate or refute a control region phylogeny (Moore 1995). Differences between the mitochondrial and nuclear trees could represent (1) differences in the accumulation of autapomorphies, (2) differential effects of lineage sorting since the introgression event (since mitochondrial loci are four times more susceptible than nuclear genes to genetic drift; Wilson et al. 1985; Moore 1995; but see Hoelzer 1997), or (3) the effects of sex-biased migration or hybridization (Palumbi and Baker 1994). In the pres-
ent study, the phylogeny for the nuclear homolog was less well resolved than was the phylogeny for the control region, which is consistent with male-biased dispersal or retention of ancient polymorphism; however, the poorer resolution is much more likely a reflection of slower rates of divergence in nuclear loci. Since so little divergence has occurred among nuclear sequences despite the antiquity of the introgression event, the nuclear homolog is not phylogenetically useful for analysis of population genetics of guillemots (and is unlikely to be for other closely related taxa). However, further investigation of the nature and extent of intraspecific variation in such sequences would provide insight into the mechanisms by which they evolve.

Control Region Evolution

As in the turnstone and dunlin (Wenink, Baker, and Tilanus 1994), fringilline finches and greenfinch Carduelis chloris (Marshall and Baker 1997), rat (Rattus norvegicus), human (Homo sapiens; Brown et al. 1986), and teleost fish (Lee et al. 1995), the central part of the control region is conserved. This conserved sequence area (thought to include the origin of heavy-strand DNA synthesis in mitochondria; Clayton 1992) is flanked by hypervariable regions at the 5′ and 3′ ends of the locus. Unfortunately, we were unable to obtain complete sequence at the 5′ end for all individuals, thereby lessening its phylogenetic utility for this study.

Attempts to resolve the number of A(n)C repeats in the 3′ end of the control region were unsuccessful. Because mitochondrial heteroplasmy for tandem repeats has been reported in mammals (evening bats, Nycticebus humeralis [Wilkinson and Chapman 1991]; shrews, genus Sorex [Stewart and Baker 1994]; elephant seals, Mirounga spp. [Hoezel, Hancock, and Dover 1993]), sturgeon (Acipenser spp.; Brown et al. 1996), and birds (turnstones and dunlins [Wenink, Baker and Tilanus 1994]; Charadriformes [Berg, Moum, and Johansen 1995]; common cuckoos, Cuculus canorus [Gibbs, de L. Brooke, and Davies 1996]), we believe that multiple amplification products result from multiple copies in the guillemot mitochondria rather than from a technical artifact.

Percentages of sequence divergence for interspecific comparisons vary among parts of the control region, although for all parts, spectacled and pigeon guillemots are most similar, and spectacled guillemots are more similar to black guillemots than are pigeon guillemots. Percent sequence divergence among species over the entire control region is similar to that of the only other study of avian congeneric interspecific control region divergence that we are aware of: Marshall and Baker (1997) found blue (Fringilla teydea) and common chaffinch (F. coelebs) sequences to be 7.8% divergent.

Mean pairwise sequence divergence is lower within guillemot species than within other species of birds that have been studied (Quinn 1992; Wenink, Baker and Tilanus 1994). When only the hypervariable part I is considered, black guillemots on average are 0.61% (SD = 0.57%) divergent from each other, pigeon guillemots are 0.74% (SD = 0.44%) divergent, and spectacled guillemots are 1.3% divergent (only one comparison available). These values are much lower than the 6.7% reported for snow geese (Quinn 1992) and the 8.0% reported for a subspecies of grey-crowned babbler (Pomatostomus temporalis temporalis; Edwards 1993) for part I. For the overall control region, black guillemots are 0.28% (SD = 0.20%) divergent, pigeon guillemots are 0.29% (SD = 0.14%) divergent, and spectacled guillemots are 1.2% divergent from each other. These values are lower than or similar to comparisons in turnstones (0.9%), dunlins (3.7%; Wenink, Baker and Tilanus 1994), and common chaffinches (0.4%–4.5%; Marshall and Baker 1997).

Phylogenetic Relationships Among Guillemots

In phylogenetic reconstructions of mitochondrial and nuclear sequences, black guillemots formed a monophyletic clade separate from pigeon and spectacled guillemots. This conclusion is consistent with other molecular phylogenies (Friesen, Baker, and Piatt 1996) and with the most plausible biogeographic scenario for these species, as follows. Storer (1952) suggested that Cephus originated in either the Atlantic or the Pacific Ocean, then, in an early Pliocene interglacial (ca. 5 MYA) spread along the north coast of Siberia to the other ocean. The fossil record suggests that guillemots were already present in the North Pacific 8–13 MYA (Warheit 1992); therefore, either the interoceanic dispersal of guillemot ancestors must have occurred earlier than was suggested by Storer (1952), or guillemots originated in the Pacific Ocean. Storer (1952) speculated that in a glacial period of the late Pliocene/early Pleistocene (ca. 1.5 MYA), the spectacled guillemot–pigeon guillemot ancestor would have diverged in the newly isolated Sea of Okhotsk. With decreasing temperatures in the late Tertiary, guillemot populations in both oceans were fragmented and subsequently diverged in isolation and gave rise to contemporary subspecies.

Phylogenetic relationships among control region sequences of the different subspecies differ across reconstruction methods and are incongruent with subspecies boundaries as defined morphologically. Such incongruence between morphological and molecular data may result either from gene flow or from retained ancestral polymorphism. A population-level survey of sequence variability suggests strong geographic partitioning of genotypes and low levels of gene flow among populations (unpublished data). The lack of resolution of genotypes of subspecies is more likely due to retention of polymorphisms from the progenitors of the modern subspecies. When an ancestral population is genetically polymorphic, descendant taxa inherit selectively neutral genotypes at random. Over time, the process of lineage sorting (“self-pruning” of phylogenetic trees through lineage extinction) will occur and leave taxa reciprocally monophyletic. Although the rate of lineage sorting fluctuates with demographic parameters such as long-term effective population size ($N_e$) and rate of population growth, populations tend to evolve reciprocal monophyly within $4N_e$ generations of isolation (Avise, Neigel, & Guries 1987).
If subspecies evolved during the last Pleistocene glaciation as suggested by Storer (1952; ca. 100,000 years ago), lineage sorting could only have occurred if $N_e$ of each subspecies was less than 5,000 females (100,000 years/4–5 years/generation for guillemots). Despite the potential for population bottlenecks during glaciation of the Northern Hemisphere (Warheit 1992; Hewitt 1996), current regional census sizes of black guillemot subspecies are much greater than this value (Finland–20,000 pairs, Norway–15,000 pairs, Svalbard–15,000 pairs, Iceland–50,000 pairs, eastern low Canadian Arctic–10,000 pairs [Nettleship and Evans 1985]). Therefore, the most likely explanation for the lack of phylogenetic resolution in this study is simply the recent divergence of the taxa.

Acknowledgments

Many thanks to the following people who helped with field collections or provided tissue samples: G. Divoky, S. Doran, A. Gaston, M. Hario, L. Hayes, M. Hipfner, K. Hobson, J. Hodder, H. Jones, M. Kassera, A. Petersen, J. Piatt, A. Prichard, S. Speckman, and H. Walsh. Thanks also to the Nature Conservancy (Maine) and La Société Duvettenor Léée (PQ, Canada) for granting access to collect samples in conservation areas. H. Walsh and T. Birt provided unpublished primer sequences. T. Birt, P. Boag, B. Congdon, C. Eckert, S. Lougheed, R. Montgomery, J. Szeicz, the Queen’s molecular ecology group, C. Moritz, and two anonymous reviewers provided valuable discussion. This study was supported by grants from the Natural Sciences and Engineering Research Council (VLF), the Northern Studies Training Program (MGK), and the Canadian Society of Ornithologists (MGK).

LITERATURE CITED


Quinn, T. W. 1992. The genetic legacy of mother goose—phylogeographic patterns of lesser snow goose Chen cae-


Craig Moritz, reviewing editor
Accepted October 6, 1997