Hereditas 119: 245-252 (1993)

Cytochrome *b* nucleotide sequence variation among the Atlantic Alcidae

V. L. FRIESEN[1]*, W. A. MONTEVECCHI[1] and W. S. DAVIDSON

[2]*Departments of Biology*[1]*, Psychology*[1] *and Biochemistry*[2]*, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X9, Canada*

FRIESEN, V. L., MONTEVECCHI, W. A. and DAVIDSON, W. S. 1993. Cytochrome *b* nucleotide sequence variation among the Atlantic Alcidae. — *Hereditas 119*: 245-252. Lund, Sweden. ISSN 0018-0661. Received January 5, 1993. Accepted July 1, 1993.

Analysis of cytochrome *b* nucleotide sequences of the six extant species of Atlantic alcids and a gull revealed an excess of adenines and cytosines and a deficit of guanines at silent sites on the coding strand. Phylogenetic analyses grouped the sequences of the common (*Uria aalge*) and Brünnich's (*U. lomvia*) guillemots, followed by the razorbill (*Alca torda*) and little auk (*Alle alle*). The black guillemot (*Cepphus grylle*) sequence formed a sister taxon, and the puffin (*Fratercula arctica*) fell outside the other alcids. Phylogenetic comparisons of substitutions indicated that mutabilities of bases did not differ, but that C was much more likely to be incorporated than was G. Imbalances in base composition appear to result from a strand bias in replication errors, which may result from selection on secondary RNA structure and/or the energetics of codon-anticodon interactions.

V. L. Friesen, Department of Ornithology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada

Imbalances in nucleotide composition of mitochondrial DNA (mtDNA) have been reported from a wide variety of vertebrates: Adenine is generally over-represented and guanine generally under-represented on the light strand (e.g., ANDERSON et al. 1981, 1982; AQUADRO and GREENBERG 1983; ROE et al. 1985; GADALETA et al. 1989; KOCHER et al. 1989; DESJARDINS and MORAIS 1990; ÁRNASON et al. 1991; BARTLETT and DAVIDSON 1991; CARR and MARSHALL 1991; EDWARDS et al. 1991; IRWIN et al. 1991; MCVEIGH et al. 1991; ÁRNASON and JOHNSSON 1992; BIRT et al. 1992; TZENG et al. 1992). Elucidation of the causes of these biases is critical both for understanding mechanisms of molecular evolution and for improving the resolution of phylogenetic analyses (e.g., KNIGHT and MINDELL 1993). Recent advances in molecular phylogenetics may provide new insights into this bias: By mapping states (nucleotides) of variable characters (nucleotide positions) onto known phylogenies, directions of substitutions may be inferred and biases in mutation directions may be investigated (e.g., AQUADRO and GREENBERG 1983; WU and MAEDA 1987; THOMAS and BECKENBACH 1989; KNIGHT and MINDELL 1993).

The avian family Alcidae comprises a distinct assemblage of Northern Hemisphere, pursuit-diving seabirds. Phylogenetic relationships among the alcids have been examined using behaviour, morphology (STRAUCH 1985), allozymes (M. K. PECK, unpubl. data) and nuclear DNA (SIBLEY and AHLQUIST 1990). Although relationships among several species are still uncertain, associations among the Atlantic species are fairly clear. In the present paper, nucleotide sequences of part of the mitochondrial cytochrome *b* gene are compared among the six extant species of Atlantic alcids to gain insight into mutational and selective forces acting on nucleotide composition.

Materials and methods

Muscle or blood was collected from one specimen each of common (*Uria aalge*) and Brünnich's (*U. lomvia*) guillemots, razorbill (*Alca torda*), little auk (*Alle alle*), black guillemot (*Cepphus grylle*), and puffin (*Fratercula arctica*; all sampled in Newfoundland). A herring gull (*Larus argentatus*; family Laridae) was used as an outgroup taxon. Nucleotide sequences of both heavy and light strands of a 307 base pair (bp) segment of cytochrome *b* were obtained as described (BIRT et al.

* Present address: Department of Ornithology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada

```
Amino Acid                  F   G   S   L   L   G   I   C   L     9
Common Guillemot        C TTC GGA TCC TTA CTA GGT ATC TGC CTA    28
Brünnich's Guillemot    . ... ... ... ... ... ..C ... ... ...    28
Razorbill               . ... ... ... C.. ... ..C ... ..T T..    28
Little Auk              . ... ... ... C.C ... ... ..T ... ..G    28
Black Guillemot         . ..T ... ... C.. ... ..C ..T ... ...    28
Puffin                  . ... ... ... C.. ..G ... ..T ... ...    28
Herring Gull            . ... ... ... C.. ... ..C ..T ... ...    28

AA            L   T   Q   I   L   T   G   L   L   L   A   T      21
CG          CTA ACA CAA ATC CTC ACA GGA CTA CTC CTA GCC ACG      64
BrG         ... ... ... ... ... ... ... ..C ..A T.. ... ..A      64
R           ... ... ... ... ..T ... ... ..C ..A ... ... ..A      64
LA          ... ... ... ... ..T ... ... ..C ..A ... ... ..A      64
BlG         ... ... ... ... ... ... ... ..T ..A ... ... ..A      64
P           AC. ... ... ... ... ... ... ..C T.A ... ... .T.      64
HG          ... ... ... ... ..A ... ... ..C ..G ... ..T .TA      64

AA            H   Y   T   A   D   T   T   L   A   F   S   S      33
CG          CAC TAC ACT GCA GAC ACA ACA CTA GCC TTT TCA TCC     100
BrG         ... ... ... ... ... ... ... ... ... ... ... ...     100
R           ..T ... ... ... ..T ... ... ... ..T ..C ... ...     100
LA          ..T ... ... ... ... ... ... ... ... ..C ..G ...     100
BlG         ..T ... ..C ... ... ... ..C ... ... ..C ... ...     100
P           ... ... ..C ... ..T ... ..C ... ... ..C ... ...     100
HG          ..T ... ..C ... ... ... ..C ... ... ..C ... ...     100

AA            V   A   H   T   C   R   N   V   Q   Y   G   W      45
CG          GTC GCC CAC ACA TGT CGA AAC GTA CAA TAT GGC TGA     136
BrG         ..T ... ... ... ... ... ... ..T ... ... ... ...     136
R           ..T ... ... ... ... ... ... ... ... ... ..T ...     136
LA          ... ... ..T ... ... ... ... ..T ... ..C ... ...     136
BlG         ..T ... ... ..C ..C ... ..T ..T ... ... ... ...     136
P           ..T ... ... ... ..C ... ... ..C ... ... ... ...     136
HG          ... ... ... ... ... ... ... ... ... ... ... ...     136

AA            L   I   R   N   L   H   A   N   G   A   S   F      57
CG          CTA ATT CGA AAC CTC CAT GCA AAC GGA GCA TCC TTC     172
BrG         ..G ... ... ... ... ... ... ... ... ... ... ...     172
R           ... ... ..G ... ... ... ... ... ... ... ... ...     172
LA          ... ... ... ... ... ... ... ... ... ... ... ...     172
BlG         ... ..C ... ... ... ... ... ... ... ... ..A ...     172
P           T.. ..C ... ... ..T ..C ... ... ... ... ..A ..T     172
HG          ... ..C ... ... ... ..C ... ... ... ..G ..A ...     172
```

Fig. 1. (to be continued)

1992). Parsimony analysis was conducted using the exhaustive search algorithm of PAUP (SWOFFORD 1990). Phenetic analysis assuming equal rates of substitutions among lineages was conducted on total numbers of substitutions, using UPGMA (unweighted pair-group method, arithmetic averages) on NTSYS-pc (ROHLF 1988). A phylogenetic hypothesis was derived from behavioural, morphological (STRAUCH 1985), allozyme (M. K. PECK unpubl. data), DNA-DNA hybridization (SIBLEY

AA	F	F	I	C	I	Y	L	H	I	G	R	G	69
CG	TTC	TTC	ATC	TGC	ATT	TAC	CTC	CAC	ATC	GGA	CGA	GGA	208
BrG	...	...	...	...	...	...	...	...	..T	...	...	...	208
R	...	...	...	..T	..C	...	...	...	..T	...	...	...	208
LA	...	...	...	...	..C	..T	...	...	...	...	...	...	208
BlG	...	...	...	...	..C	...	...	...	..T	...	...	...	208
P	...	...	...	...	..C	...	...	...	..T	...	...	...	208
HG	...	..T	..T	..T	...	...	..A	...	...	...	...	...	208

AA	F	Y	Y	G	S	Y	L	N	K	E	T	W	81
CG	TTC	TAC	TAC	GGC	TCA	TAC	CTG	AAC	AAA	GAA	ACC	TGA	244
BrG	...	...	...	...	..G	...	..A	...	...	...	...	...	244
R	...	...	...	...	...	...	..A	...	...	..G	...	...	244
LA	...	...	...	...	...	...	..A	...	...	...	...	...	244
BlG	...	...	...	...	...	...	T.A	...	...	...	...	...	244
P	..T	...	...	..T	..C	...	..A	...	...	...	...	...	244
HG	...	...	..T	...	...	...	..C	**T**.T	...	...	...	...	244

AA	N	T	G	V	I	L	L	L	A	L	M	A	93
CG	AAC	ACA	GGT	GTC	ATC	CTC	CTA	CTA	GCC	CTA	ATG	GCA	280
BrG	..T	...	...	...	..T	...	...	...	...	...	..A	...	280
R	...	...	...	..T	..T	..T	...	...	...	..T	..A	...	280
LA	...	...	..C	**A**..	..T	..T	...	...	...	...	..A	...	280
BlG	...	...	..C	**A**..	...	...	..T	T..	**A**..	...	..A	...	280
P	...	...	..C	...	..T	...	...	..G	...	...	..A	...	280
HG	..T	...	..A	...	..T	...	..C	T..	**A**..	...	..A	...	280

AA	T	A	F	V	G	Y	V	L	P	102
CG	ACC	GCT	TTC	GTA	GGA	TAT	GTC	CTC	CCA	307
BrG	...	..C	...	..G	..T	...	..T	..A	...	307
R	..T	..C	...	...	..G	...	...	..A	...	307
LA	..T	...	...	...	...	..C	...	..T	...	307
BlG	..A	..C	...	...	...	...	...	..A	...	307
P	..T	..C	...	...	...	...	...	..G	...	307
HG	..T	..C	...	...	...	...	...	T.G	...	307

Fig. 1. (continued)

Fig. 1. Coding strand sequences for a fragment of cytochrome *b* for six alcid species and a gull. Dots indicate identity with the common guillemot sequence. Nucleotide substitutions resulting in amino acid substitutions are highlighted in bold. Aa = inferred amino acid sequence of the common guillemot; CG = common guillemot: BrG = Brünnich's guillemot; R = razorbill; LA = little auk; BlG = black guillemot; P = puffin; HG = herring gull.

and AHLQUIST 1990), and cytochrome *b* data. Silent substitutions were mapped onto this phylogeny using PAUP, and directions of silent substitutions for which ancestral states could be inferred were tallied. Availabilities of bases for substitutions were estimated from numbers of bases at silent sites in ancestral sequences, also determined using PAUP.

Results and discussion

Variability of nucleotide and amino acid sequences

Twenty-six percent of nucleotide sites (79 of 307) were variable among the seven species examined (Fig. 1). Most mutations occurred at third positions of codons, and few occurred at second positions: 63 % of third positions were variable,

whereas only 12 % of first and 2 % of second positions varied (Fig. 1). Standard deviations in numbers of bases at the three positions showed a similar pattern, being greatest at third and lowest at second positions (Table 1). These results are consistent with the observation that most first and all second position substitutions produce amino acid replacements, whereas third positions are freer to vary because of the degeneracy of the genetic code. Furthermore, substitutions at second positions cause more radical amino acid replacements than do those at first positions, so are more likely to be eliminated by selection (BROWN 1985; LI et al. 1985). Most (eight [67 %] of 12) first position substitutions involved leucine codons and did not result in amino acid changes. Only five amino acid positions varied (Fig. 1; Table 2): Four involved changes in functional groups, but all have been reported before and/or occurred at sites not thought to be essential for enzymatic function of cytochrome *b* (HOWELL and GILBERT 1988; KOCHER et al. 1989; WEIR 1990; IRWIN et al. 1991). Transitional substitutions outnumbered transversions for all comparisons (Table 2; Fig. 1); a similar bias has been reported in many sequence analyses of mtDNA (e.g., WU and MAEDA 1987; KOCHER et al. 1989; THOMAS and BECKENBACH 1989; BARTLETT and DAVIDSON 1991; CARR and MARSHALL 1991; EDWARDS et al. 1991; IRWIN et al. 1991; MCVEIGH et al. 1991; BIRT et al. 1992; KNIGHT and MINDELL 1993).

Table 1. Base usage (mean number ± *SD*) by codon position within a segment of the cytochrome *b* gene for six alcid species. Percents are in parentheses

Base	Codon position		
	1	2	3
A	26 ± 0.8 (25 %)	21 ± 0.0 (21 %)	40 ± 1.0 (39 %)
C	26 ± 1.0 (25 %)	25 ± 0.0 (25 %)	44 ± 3.9 (43 %)
G	25 ± 0.8 (25 %)	17 ± 0.0 (17 %)	3 ± 1.0 (3 %)
T	25 ± 0.8 (25 %)	39 ± 0.0 (38 %)	16 ± 3.8 (15 %)

Table 2. Observed numbers of amino acid/nucleotide substitutions (above diagonal) and transitions/transversions (below diagonal) in a segment of cytochrome *b* among six alcid species and a gull. Abbreviations as in Fig. 1

	CG	BrG	R	LS	BlG	P	HG
Common guillemot	–	0/19	0/29	0/25	2/31	3/35	1/36
Brünnich's guillemot	14/5	–	0/25	0/29	2/28	3/35	1/38
Razorbill	25/4	22/3	–	0/27	2/30	2/37	2/39
Little auk	21/4	26/3	22/5	–	0/29	2/36	2/38
Black guillemot	22/9	22/6	23/7	23/6	–	4/30	2/34
Puffin	27/8	30/5	31/6	30/6	25/5	–	2/38
Herring Gull	25/11	28/10	30/9	27/11	26/8	29/9	–

The four nucleotides were equally represented at first codon positions (Table 1). A slight excess of Ts occurred at second positions, and a marked imbalance in base usage occurred at third positions: A and C were heavily used, whereas G was seldom used. Similar inequalities have been reported for a wide variety of vertebrates both for cytochrome *b* (KOCHER et al. 1989; BARTLETT and DAVIDSON 1992; CARR and MARSHALL 1991; EDWARDS et al. 1991; IRWIN et al. 1991; MCVEIGH et al. 1991; BIRT et al. 1992) and for the mitochondrial genome as a whole (ANDERSON et al. 1981, 1982; ROE et al. 1985; GADALETA et al. 1989; DESJARDINS and MORAIS 1990).

Phylogenetic analyses

Cladistic analysis of 35 phylogenetically informative sites produced two shortest trees (total length = 68 steps; consistency index = 0.63; Fig. 2). Both trees grouped the sequences of the common and Brünnich's guillemots, followed by the razorbill and little auk. The black guillemot sequence formed a sister taxon, and the puffin fell outside the other alcids. Results of the phenetic analysis were similar. Neither analysis could resolve the position of the razorbill relative to the *Uria* guillemots and little auk (Fig. 2). This uncertainty probably does not result from intra-specific variability: Maximum percent sequence divergence within this segment of cytochrome *b* among 219 Brünnich's guillemots from throughout the Northern Hemisphere (2.4 %; BIRT-FRIESEN et al. 1992) was much lower than sequence divergence among species (Fig. 2; Table 2). The ambiguity may indicate that *Uria*, *Alca*, and *Alle* diverged within a short time period.

Results of both phenetic and cladistic analyses are consistent with results of comparisons of behaviour, morphology (STRAUCH 1985), allozymes (M. K. PECK, unpubl. data), and DNA-DNA hybridization (SIBLEY and AHLQUIST 1990). The phylogeny presented in Fig. 2 therefore provides a robust hypothesis for investigation of mutation rates. When silent substitutions were mapped onto this tree, inequalities were apparent in directions of mutations, and included biases in both ancestral and incorporated bases (Table 3a). The bias in ancestral bases appeared to relate to availabilities of bases for mutation. Expected num-

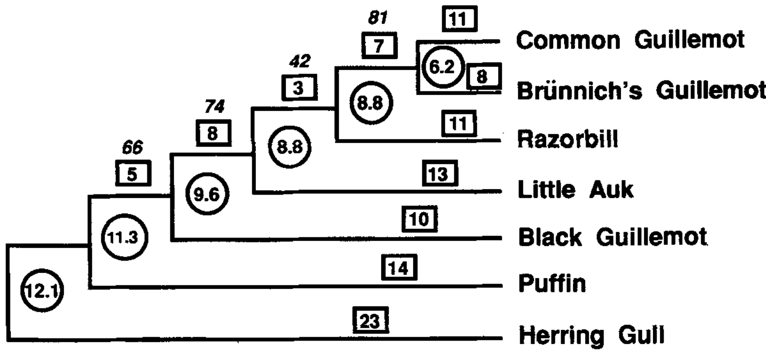

Fig. 2. Phylogenetic hypothesis for the Atlantic Alcidae. Numbers in boxes indicate minimum numbers of nucleotide substitutions within a fragment of cytochrome *b* within each lineage. Numbers in italics are the percent of times that each association occurred in 1000 bootstrap replicates, performed on PAUP. Numbers in circles indicate percent sequence divergence from UPGMA analysis of data in Table 2.

Table 3a. Incidences of silent substitutions within a segment of the cytochrome *b* gene for six alcid species. Numbers of bases available for substitution (from base compositions of ancestral sequences) are in parentheses

Ancestral base	Incorporated base				Total
	A	C	G	T	
A	–	6 (207)	15 (252)	4 (207)	25
C	2 (95)	–	0 (95)	36 (345)	38
G	1 (1)	0 (0)	–	0 (0)	1
T	1 (14)	11 (59)	0 (14)	–	12
Total	4	17	15	40	76

Table 3b. Mutabilities of bases. Observed values from Table 3a; expected values calculated from base availabilites (from Table 3a) (Total = 76)

Ancestral base	Observed number	Expected number	χ^{2*} components
A	25	29.15	0.46
C	38	39.91	0.05
G	1	0.12	1.20
T	12	6.82	3.21
χ^2			4.92
P			ns

* Including Yates' correction

Table 3c. Incorporation of bases. Observed values from Table 3a; expected values calculated from base availabilities and relative rates of transitions and transversions (63:13; from Table 3a) (Total = 76)

Incorporated base	Observed number	Expected number	χ^{2*} components
A	4	4.09	0.04
C	17	8.16	8.52
G	15	28.16	5.69
T	40	35.59	0.43
χ^{2*}			14.68
P			< 0.01

* Including Yates' correction

Table 3d. Incidences of transitions. Observed values from Table 3a; expected values calculated from base availabilities (from Table 3a) (Total = 63)

Transition	Observed number	Expected number	χ^{2*} components
A > G	15	24.16	3.10
C > T	36	33.08	0.18
G > A	1	0.10	1.70
T > C	11	5.66	4.14
χ^2			9.12
P			<0.05

* Including Yates' correction

bers of substitutions after correction for biases in ancestral base composition, accounted for most of the apparent inequality in mutability (Table 3b); i.e., all bases were equally likely to mutate. However, biases in the incorporated base were apparent

even when expected numbers were corrected for availabilities of ancestral bases and differences in incidences of transitions and transversions (Table 3c): C was more likely to be incorporated than was G. Similar results have been reported in other studies (e.g., AQUADRO and GREENBERG 1983; THOMAS and BECKENBACH 1989; KNIGHT and MINDELL 1993).

Explanation of biases in base usage

Several explanations for the observed bias in base composition have been proposed (e.g., LI et al. 1985; WU and MAEDA 1987; KNIGHT and MINDELL 1993).

Biased amino acid composition. — The bias towards Ts at second positions (Table 1) correlates with a high representation of leucines (Fig. 1; Table 4): Leucine is encoded as CTN or TTR (where N represents any nucleotide and R represents either purine) and comprises 18 % of amino acids in this segment in the species examined. However, inequalities in amino acid representation do not explain the imbalance in base composition at silent sites, since codon usage was also biased (Table 4): At two-fold degenerate sites involving purines, A was used 13 times more than G, and at two-fold degenerate sites involving pyrimidines, C was used 4 times more than T. An imbalance also occurred at four-fold degenerate sites, where the ratio of A:C:G:T was 14:9:1:3. Furthermore, biases in amino acid composition may be a result rather than a cause of biased base composition to some extent.

Table 4. Codon usage within a segment of cytochrome *b* among six alcid species. The first number for each codon represents the number of times the codon is used at variable sites; the second number represents the total number of times the codon is used. Codons that match anticodons for chicken mtDNA (DESJARDINS and MORAIS 1990) are highlighted in bold

Codon	Times used	Codon	Times used	Codon	Times used
Alanine		Glycine		Proline	
GCA	0/24	**GGA**	3/33	**CCA**	0/6
GCC	9/26	GGC	16/16	CCC	0/0
GCG	0/0	GGG	2/2	CCG	0/0
GCT	3/3	GGT	9/9	CCT	0/0
Arginine		Histidine		Serine	
CGA	5/17	**CAC**	9/15	**TCA**	11/11
CGC	0/0	CAT	9/9	TCC	5/17
CGG	1/1			TCG	3/3
CGT	0/0	Isoleucine		TCT	0/0
		ATC	13/27	**AGC**	0/0
Asparagine		ATT	17/17	AGT	0/0
AAC	4/4				
AAT	2/2	Leucine		Threonine	
		CTA	48/59	**ACA**	14/39
Aspartic Acid		CTC	20/26	ACC	7/14
GAC	10/28	CTG	6/6	ACG	1/1
GAT	2/2	CTT	9/9	ACT	7/7
		TTA	7/7		
Cysteine		TTG	0/0	Tryptophan	
TGC	12/12			**TGA**	12/12
TGT	6/6	Lysine		TGG	0/0
		AAA	0/6		
Glutamic Acid		AAG	0/0	Tyrosine	
GAA	5/5			**TAC**	7/31
GAG	1/1	Methionine		TAT	11/11
		ATA	5/5		
Glutamine		**ATG**	1/2	Valine	
CAA	0/12			**GTA**	7/7
CAG	0/0	Phenylalanine		GTC	11/11
		TTC	19/37	GTG	1/1
		TTT	5/5	GTT	9/9

Avoidance of pretermination codons. — CLARKE (1970) suggested that codons that differ by one substitution from termination codons (AGA, AGG, TAA and TAG in mtDNA) may be avoided to minimize potential mutational damage. Although many pretermination codons, such as AGC and AGT, were avoided in the present data set, several, such as ACA and GGA, were heavily used.

Codon-anticodon interactions. — LI et al. (1985) argued that the bias in codon usage in nuclear DNA may result from tRNA availabilities. This explanation does not apply to mtDNA, since the mitochondrial genomes of vertebrates encode only 22 transfer RNAs (tRNAs): Two each for serine and leucine and one for each other amino acid (ANDERSON et al. 1981, 1982; ROE et al 1985; GADALETA et al. 1989; DESJARDINS and MORAIS 1990; ÁRNASON et al. 1991; ÁRNASON and JOHNSSON 1992; TZENG et al. 1992; Table 4). Biases in base usage may, however, result from the energetics of codon-anticodon interactions, since the codon that matches the tRNA will form the most stable bond during translation. In the present results, codons that matched anticodons for the tRNAs of the chicken (*Gallus domesticus*; DESJARDINS and MORAIS 1990; Table 4) were used for most (17 of 22) tRNAs. Almost all anticodons in chicken mtDNA end with U or G, possibly because these have the greatest flexibility for base pairing (ANDERSON et al. 1981); anticodon usage may therefore explain the bias towards As and Cs at third positions of codons. However, it does not explain the bias against Gs in regions that do not code for proteins, such as the control region (AQUADRO and GREENBERG 1983) or ribosomal RNAs (rRNAs; KNIGHT and MINDELL 1993).

Secondary structure and codon context. — Little is known about secondary structure or codon context (interactions between neighbouring bases) in mtDNA (LI et al. 1985). Secondary structure is clearly important in tRNA and rRNA genes and in parts of the control region (e.g., ROE et al. 1985; BROWN et al. 1986; GADALETA et al. 1989). BROWN (1983) observed that the bias against Gs in protein-coding genes is strongest in homeotherms, and suggested that it may result from a tendency of guanines to form secondary structure in messenger RNA (mRNA). Selection for intermediate stability in the secondary structure of mRNA may also favour random distributions of weak (A-T) and strong (C-G) bonds, which may result in nonrandom codon usage (LI et al. 1985). Accordingly, the distribution of weakly and strongly bonding bases (Cs and Gs versus As and Ts) within the cytochrome *b* sequence of the common guillemot is distinctly random (Fig. 1; runs test, $P = 0.98$). However, this explanation for biases in base composition is not completely satisfactory since ND6, which is coded on the heavy (instead of the light) strand of the mitochondrial genome, has an excess of Gs and Ts and deficit of Cs and As on the coding strand (ANDERSON et al. 1981, 1982; AQUADRO and GREENBERG 1983; ROE et al. 1985; GADALETA et al. 1989; DESJARDINS and MORAIS 1990; ÁRNASON et al. 1991; ÁRNASON and JOHNSSON 1992; TZENG et al. 1992).

Non-random mutation. — Biases in base composition may result from differences in rates of different mutations. Mutations can occur through a variety of mechanisms, most of which produce transitions (LINDAHL 1993). If differences in base composition result from simple mutational biases, then each type of substitution should be equally likely on either strand. Thus A to G transitions on the coding strand should be as common as A to G transitions on the noncoding strand or T to C transitions on the coding strand. However, T to C transitions were much more frequent than A to G transitions in the present data, even when expected numbers of mutations were corrected for base availability (Table 3d). Similar results have been reported in other studies (e.g., AQUADRO and GREENBERG 1983; WU and MAEDA 1987). Furthermore, despite the paucity of Gs on the coding strand, Gs were common on the noncoding strand (Table 3a). This suggests a strand bias in mutation rates, rather than a simple mutational bias towards As or Cs over Gs.

Strand bias in replication errors. — WU and MAEDA (1987) argued that differences in the replication mechanisms between the two strands of DNA may result in differences in mutation rates. This hypothesis is consistent with observed strand biases in both base usage and mutation rates in mtDNA, including biases for protein-coding genes, RNA genes and the control region (above). It is also easily reconciled with observations on the energetics of codon-anticodon interactions and secondary structure (above). For example, if an excess of Gs results in unfavourable secondary structure in RNA and/or unfavourable bonds at 'wobble' positions of anticodons, selection may have favoured a replication mechanism that reduces the incorporation of Gs on the light strand, as well as anticodons containing Us or Gs at third positions. (Unfavourable codon/anticodon interactions due to a bias towards Gs on the coding strand of ND6 may not be sufficient to overcome the bias in replication errors.) This hypothesis is especially attractive because it does not require the invocation of selection on individual bases to explain biases in base usage. Mechanisms of replication of mtDNA may warrant further investigation.

Acknowledgements. — We are grateful to Jay Pitocchelli and Pierre Ryan for help with sample collections. Allan Baker, Sylvia Bartlett, Tim Birt, Steve Carr and Dawn Marshall assisted with laboratory and/or computer analyses. Allan Baker, Tim Birt, Steve Carr, Keith Egger and two anonymous reviewers provided valuable comments on the manuscript. Mark Peck generously shared unpublished data. Funding was provided by Natural Sciences and Engineering Research Council operating grants to W. A. M. and W. S. D. and a postgraduate scholarship to V. L. F.

References

ANDERSON, S., BANKIER, A. T., BARRELL, B. G., DE BRUIJN, M. H. L., COULSON, A. R., DROUIN, J., EPERON, I. C., NIERLICH, D. P., ROE, B. A., SANGER, F., SCHREIER, P. H., SMITH, A. J. H., STADEN, R. and YOUNG, I. G. 1981. Sequence and organization of the human mitochondrial genome. — *Nature 290*: 457–465

ANDERSON, S., DE BRUIJN, M. H. L., COULSON, A. R., EPERON, I. C., SANGER, F. and YOUNG, I. G. 1982. Complete sequence of bovine mitochondrial DNA. — *J. Mol. Biol. 156*: 683–717

AQUADRO, C. F. and GREENBERG, B. D. 1983. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. — *Genetics 103*: 287–312

ÁRNASON, U. and JOHNSSON, E. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. — *J. Mol. Evol. 34*: 493–505

ÁRNASON, Ú., GULLBERG, A. and WIDEGREN, B. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. — *J. Mol. Evol. 33*: 556–568

BARTLETT, S. E. and DAVIDSON, W. S. 1991. Identification of *Thynnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial DNA. — *Can. J. Fish. Aquat. Sci. 48*: 309–317

BIRT, T. P., BIRT-FRIESEN, V. L., GREEN, J. M., MONTEVECCHI, W. A. and DAVIDSON, W. S. 1992. Cytochrome-*b* sequence variation among parrots. — *Hereditas 117*: 67–72

BIRT-FRIESEN, V. L., MONTEVECCHI, W. A., GASTON, A. J. and DAVIDSON, W. S. 1992. Genetic structure of thick-billed murre (*Uria lomvia*) populations examined using direct sequence analysis of amplified DNA. — *Evolution 46*: 267–272

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

BROWN, W. M. 1983. Evolution of animal mitochondrial DNA. — In: *Evolution of Genes and Proteins* (*eds* M. NEI *and* R. K. KOEHN), *Sinauer, Sunderland, Mass*, p. 62–88

BROWN, W. M. 1985. The mitochondrial genome of animals. — In: *Molecular Evolutionary Genetics* (*ed* R. J. MACINTYRE), *Plenum, London, England*, p. 95–130

CARR, S. M. and MARSHALL, H. D. 1991. Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. — *Can. J. Fish. Aquat. Sci. 48*: 48–52

CLARKE, B. C. 1970. Darwinian evolution of proteins. — *Science 168*: 1009–1011

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

EDWARDS, S. V., ARCTANDER, P. and WILSON, A. C. 1991. Mitochondrial resolution of a deep branch in the geneological tree for perching birds. — *Proc. R. Soc. Lond. B 243*: 99–107

GADALETA, G., PEPE, G., DE CANDIA, G., QUAGLIARIELLO, C., SBISÁ, E. and SACCONE, C. 1989. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. — *J. Mol. Evol. 28*: 497–516

HOWELL, N. and GILBERT, K. 1988. Mutational analysis of the mouse mitochondrial cytochrome *b* gene. — *J. Mol. Biol. 203*: 607–618

IRWIN, D. M., KOCHER, T. D. and WILSON, A. C. 1991. Evolution of the cytochrome *b* gene of mammals. — *J. Mol. Evol. 32*: 128–144

KNIGHT, A. and MINDELL, D. P. 1993. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of Fea's viper. — *Syst. Biol. 42*: 18–31

KOCHER, T. D., THOMAS, W. K., MEYER, A., EDWARDS, S. V., PÄÄBO, S., VILLABLANCA, F. X. and WILSON, A. C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. — *Proc. Natl. Acad. Sci. USA 86*: 6196–6200

LI, W.-H., LUO, C.-C. and WU, C.-I. 1985. Evolution of DNA sequences. — In: *Molecular Evolutionary Genetics* (*ed* R. J. MACINTYRE), *Plenum, London, England*, p. 1–94

LINDAHL, T. 1993. Instability and decay of the primary structure of DNA. — *Nature 362*: 709–715

MCVEIGH, H. P., BARTLETT, S. E. and DAVIDSON, W. S. 1991. Polymerase chain reaction/direct sequence analysis of the cytochrome *b* gene in *Salmo salar*. — *Aquaculture 95*: 225–233

ROE, B. A., MA, D.-P., WILSON, R. K. and WONG, J. F.-H. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. — *J. Biol. Chem. 260*: 9759–9774

ROHLF, F. J. 1988. NTSYS-pc: Numerical taxonomy and Multivariate Analysis System, version 1.50. — *Exeter, Setauket, New York*

SIBLEY, C. G. and AHLQUIST, J. E. 1990. Phylogeny and Classification of Birds. — *Yale University Press, New Haven, Connecticut*

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

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.0s. — *Illnois Natural History Survey, Champaigne, IL*

THOMAS, W. K. and BECKENBACH, A. T. 1989. Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. — *J. Mol. Evol. 29*: 233–245

TZENG, C.-S., HUI, C.-F., SHEN, S.-C. and HUANG, P. C. 1992. The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. — *Nucleic Acids Res. 20*: 4853–4858

WEIR, B. S. 1990. Genetic Data Analysis. — *Sinauer, Sunderland, Mass.*

WU, C.-I. and MAEDA, N. 1987. Inequality in mutation rates of the two strands of DNA. — *Nature 327*: 169–170