Acta Biochimica Polonica

Vol. 51 No. 4/2004

897–905

QUARTERLY

Non-random base composition in codons of mitochondrial cytochrome *b* gene in vertebrates*

Beata Prusak[1]✉ and Tomasz Grzybowski[2]

[1]*Department of Animal Immunogenetics, Institute of Genetics and Animal Breeding, Polish Academy of Sciences, Jastrzębiec, Poland;* [2]*The Ludwik Rydygier University School of Medical Sciences, Forensic Medicine Institute, Bydgoszcz, Poland*

Received: 30 April, 2004; revised: 01 September, 2004; accepted: 07 October, 2004

Key words: cytochrome *b*, compositional bias, structure-function relationships

Cytochrome *b* is the central catalytic subunit of the quinol:cytochrome *c* oxidoreductase of complex III of the mitochondrial oxidative phosphorylation system and is essential to the viability of most eukaryotic cells. Partial cytochrome *b* gene sequences of 14 species representing mammals, birds, reptiles and amphibians are presented here including some species typical for Poland. For the analysed species a comparative analysis of the natural variation in the gene was performed. This information has been used to discuss some aspects of gene sequence – protein function relationships. Review of relevant literature indicates that similar comparisons have been made only for basic mammalian species. Moreover, there is little information about the Polish-specific species. We observed that there is a strong non-random distribution of nucleotides in the cytochrome *b* sequence in all tested species with the highest differences at the third codon position. This is also the codon position of the strongest compositional bias. Some tested species, representing distant systematic groups, showed unique base composition differing from the others. The quail, frog, python and elk prefer C over A in the light DNA strand. Species belonging to the artiodactyls stand out from the remaining ones and contain fewer pyrimidines. The observed overall rate of amino acid identity is about 61%. The region covering Q_o center as well as histidines 82 and 96 (heme ligands) are totally conserved in all tested species. Additionally, the applied method and the sequences can also be used for diagnostic species identification by veterinary and conservation agencies.

---

*GenBank Accession No: AY840094, AY840095, AY840096, AY840097, AY840098, AY840099, AY840100, AY840101, AY840102, AY840103, AY840104, AY840105, AY840106, AY840107, AY840108.

✉Address for correspondence: Beata Prusak, Department of Animal Immunogenetics, Institute of Genetics and Animal Breeding, Polish Academy of Sciences, Jastrzębiec, 05-552 Wólka Kosowska, Poland; tel.: (48 22) 756 1711; fax: (48 22) 756 1699; e-mail: b.prusak@ighz.pl

Cytochrome *b* is one of the best known proteins that make up complex III of the mitochondrial oxidative phosphorylation system (Hatefi, 1985) and is the only one encoded by the mitochondrial genome. Cytochrome *b* is a transmembrane protein consisting of eight α-helices and it is believed to contain both redox centers Q_o and Q_i (Hatefi, 1985). All eukaryotic organisms require this class of redox enzymes, and consequently cytochrome *b*, for energy conservation (Trumpower, 1990). Based on mutational studies a structural model of cytochrome *b* including the sites of electron transfer and inhibitor action has been developed (Howell & Gilbert, 1988; di Rago *et al.,* 1990). Knowledge of mitochondrial cytochrome *b* is expanding very rapidly. At present cytochrome *b* gene is also used as a tool in studies of molecular evolution (Kocher *et al.,* 1989; Montgelard *et al.,* 1997; Prusak *et al.*, 2004) and legal medicine (Bartlett & Davidson, 1992; Zehner *et al.,* 1998; Parson *et al.,* 2000). Due to higher mobility of people, products and technologies, increasing problems in food safety and extinction of many valuable species, studies focused on characterising global genetic resources and species identification based on cytochrome *b* gene are presently carried out in many countries. Cytochrome *b* gene has been completely or partially sequenced for many species of mammals, birds, reptiles, amphibians, fishes and also some invertebrates. The importance of acquiring the information about cytochrome *b* sequences for a wide range of species is also that it enables us to observe how the protein has evolved while maintaining its function. Moreover, a number of neurodegenerative disorders and metabolic myopathies have been reported in relation to mutations in cytochrome *b* and other mitochondrial genes (Lestienne *et al.,* 1999). Since mitochondrial disorders in many cases are multisystemic with two genomes and different genes being involved (Shoffner & Wallace, 1990), it seems important to describe the genetic, biochemical and functional characteristics of the enzymes making up the oxidative phosphorylation chain. In the present study we aimed to investigate some aspects of the natural variation in the cytochrome *b* gene sequence in relation to the amino-acid composition and the structure of the protein. Special attention has been paid to some species typical for Poland, being under conservation or included in the National Rare Livestock Breeds Preservation Programme, untill now not researched. Since such material requires special methodological approaches in sampling (restrictions regarding requirements for sample possessing from conserved animals), thus variable biological sources (blood, tissue, hair, blood stain) were used to demonstrate that the applied method can also be used in future studies concerning genetic characterisation of typical, rare and protected species in Poland.

MATERIALS AND METHODS

Biological material. The material covered hair, blood and soft tissues (heart, liver, muscles). Partial cytochrome *b* gene sequences of 14 species were analysed: American bison (hair), European bison (hair and liver), zebu cattle (hair), Polish Red cattle (blood) and Polish Whiteback cattle (blood) – breeds included in the National Rare Livestock Breeds Preservation Programme, goat (Polish White Improved – blood and hair), sheep (Merino breed – blood), human (hair), boar (liver), elk (blood stain), European lynx (blood), quail (heart), *Rana temporaria* (heart) and *Python molurus* (muscles).

DNA extraction and cyclic sequencing. Total genomic DNA from hair, soft tissues and blood stain was extracted according to the standard organic procedure (Wilson *et al.,* 1995). For each source of samples specific conditions of extraction were adjusted. Genomic DNA from blood samples was extracted using Wizard® Genomic DNA Purification Kit (Promega).

DNA amplification was performed using primers given by Parson *et al.* (2000). The sequences of the forward and reverse primers were modified in such a way that one oligonucleotide in a pair was extended by a universal primer sequence (–21) M13 at the 5′end. The PCR reaction was conducted in a GenAmp PCR System 9600 Thermal Cycler (AB), according to the following parameters: 94°C for 2 min (denaturation) and next 94°C for 30 s, 50°C for 45 s, 72°C for 45 s – 35 cycles. In order to improve sequencing efficiency the PCR amplification was carried out in a two step manner. First, the PCR reaction was conducted in a volume of 50 μl and the composition of the reaction mixture was as follows: 50–100 ng of genomic DNA, 200 μM of each dNTP, 1 × PCR buffer (AB), 1.5 mM $MgCl_2$, 1 μM of each primer, 1.5 U DNA Taq Gold polymerase (AB). Then a second round of amplification was made. Four microliters of 1000 × diluted PCR products were amplified under the same conditions except the quantity of primers (0.1 μM) and polymerase (1 U). The PCR products were purified through ultrafiltration using Microcon 100 microconcentrators (Amicon). The quantity and quality of products was tested in 4% NuSieve 3 : 1 agarose gel (FMC) in relation to the DNA mass ladder standard (pUC19, Ingen, Sieradz, Poland). Purified PCR products (273 base pair length) were sequenced with ABI Prism Big Dye Primer Cycle Sequencing Ready Reaction Kit (AB) according to the user's manual in a GenAmp PCR System 9600 Thermal Cycler (AB). The sequencing products were separated in a DNA sequencer ABI PRISM 377 (AB). The electrophoretic data were collected by the Data Collection v.2.1. software (AB) and analysed by the Sequencing Analysis v.3.0. software (AB).

Base and amino-acid composition. The bias in base composition was calculated according to the formula given by Irvin *et al.* (1991). The protein sequence was predicted from cytochrome *b* gene sequences using BioEdit Sequence Alignment Editor (Hall, 1999).

RESULTS AND DISCUSSION

A 271 bp fragment of cytochrome *b* gene sequence was analysed. Species included in the study comprise representatives of mammals, birds, reptiles and amphibians including some species typical for Poland or breeds included in the National Rare Livestock Breeds Preservation Programme. Until now little has been known about the genetic properties of Polish native species and breeds. Thus the most important premise in determination of primer pairs was the ability to obtain informative sequencing results for a wide range of animal groups. Preliminary results (data unpublished) showed that high sequence conservation of the chosen primers makes them suitable for mammals as well as for other vertebrate groups. The advantage is that it makes possible to compare base and amino-acid composition in main vertebrate groups using only a single primer pair. Although the present study was carried out only for a fragment of the cytochrome *b* gene, the results obtained are comparable with those concerning the complete cytochrome *b* gene, described by other authors (Irvin *et al.,* 1991). The fragment of gene chosen for the present study comprises all main parts of the cytochrome *b* protein (the inner, transmembrane and outer segments) differing in the level of interspecies variability.

For individuals of *Bison bison,* two types of sequences (type I and type II) were obtained differing in one nucleotide position: substitution T–C at position 15039 according to Anderson human reference sequence (Anderson *et al.*, 1981). This substitution is silent, with no changes in amino-acid composition. The sequences for Polish Red cattle, Polish Whiteback cattle, Zebu cattle, human, sheep, goat, boar, elk, European lynx, quail, frog and *Python molurus* were determined as single types. These sequences are shown in Fig. 1.

Although mitochondrial DNA fragments have been found in the nuclear genome of hu-

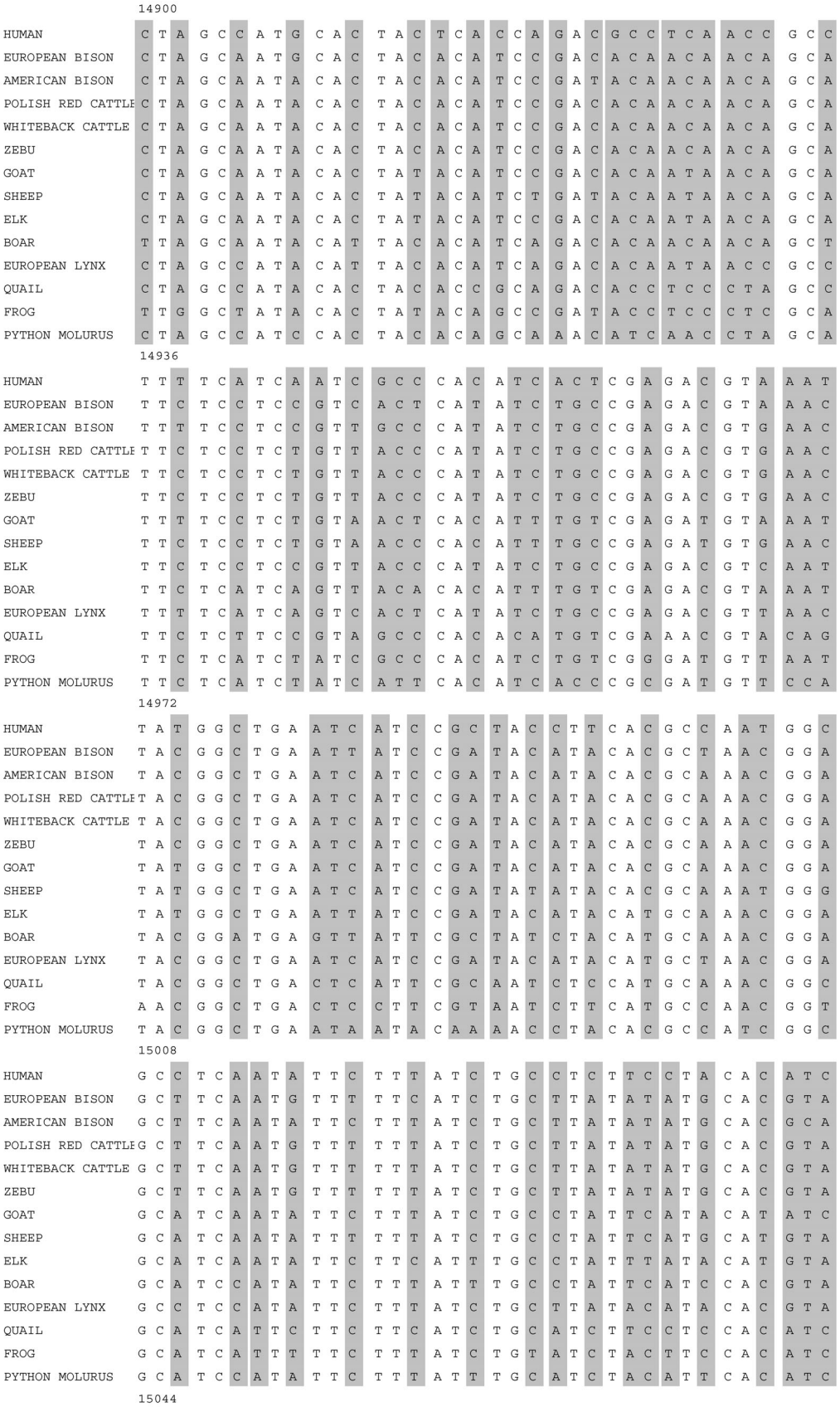

Figure 1. Sequences of analysed fragment of cytochrome *b* gene.

Shadowed positions – variable sites. Numbering is according to the human reference sequence (Anderson *et al.*, 1981).

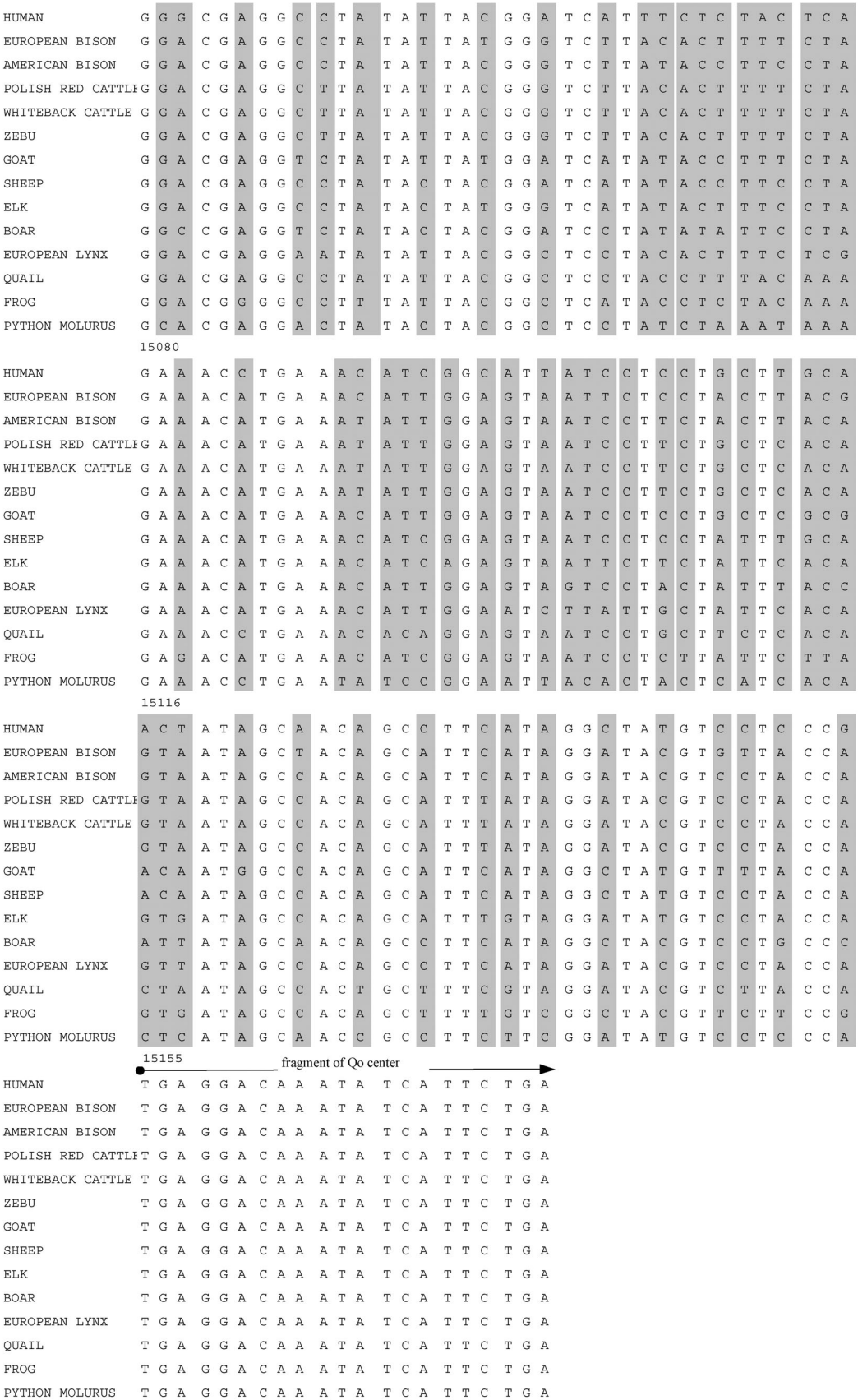

Figure 1. Continued.

mans (Fukuda *et al.,* 1985), we found no sequence behaving as a pseudogene (no sequence with two different bases detected in one position or distinctly differing in the number of substitutions (Li *et al.,* 1985)) and we concluded that all the sequences are of mitochondrial origin. A strong non-random distribution of bases in the cytochrome *b* gene fragment was found. Since many differences at the third codon positions are silent substitutions, the highest differences in base composition among species were found in this position (the highest standard deviation) (Table 1). The smallest differences occurred in the second codon position (Table 1). It is known that animal mtDNA exhibits extreme bias in base composition at silent sites, which manifests itself as a difference in composition between the two strands of mtDNA (Brown, 1985). Moreover, it has been shown that different systematic groups of mammals have different preferred bases in certain codon positions (Irvin *et al.,* 1991). There are several potential explanations of the observed differences. One of them is selection; the other is mutation pressure (Sueoka, 1988). As a consequence, the transition-transversion ratios could vary at different codon positions. This is the reason for the higher likelihood that changes at the third position are silent. It is also possible that certain nucleotide compositions could result from different base substitutions. Compositional bias was calculated according to the formula given by Irvin *et al.* (1991). The strongest bias for all species was observed at the third position, and the least at the first codon position (Table 1). Previous results dealing with nucleotide composition for complete cytochrome *b* gene reveal that codon positions differ in their composition (Irvin *et al.,* 1991). The results of our study show that the third position has very few G, the second is enriched in T, and the first is rather unbiased. This rule was observed for all representatives of the tested animal groups (mammals, birds, reptiles and amphibians). The base composition at the three codon positions of the cytochrome *b* gene is shown in Table 1. It seems that the obtained results concerning non-random distribution

Table 1. Base composition at the first, second and third position of codons.

Species	First				Second				Third			
	A	G	T	C	A	G	T	C	A	G	T	C
European bison	27.8	25.5	30.0	16.7	21.1	17.8	36.7	24.4	44.5	5.5	17.8	32.2
American bison	28.3	25.3	27.3	19.1	19.2	18.2	39.4	23.2	42.4	4.0	14.2	39.4
Polish Red cattle	27.5	25.3	30.7	16.5	20.8	18.7	36.3	24.2	45.0	5.5	16.5	33.0
Polish Whiteback cattle	27.5	25.3	30.7	16.5	20.8	18.7	36.3	24.2	45.0	5.5	16.5	33.0
Zebu	27.5	25.3	30.7	16.5	20.8	18.7	36.3	24.2	45.0	5.5	16.5	33.0
Goat	26.3	26.3	29.0	18.4	19.8	23.0	38.0	18.4	46.0	5.3	18.4	30.3
Sheep	27.0	26.0	29.0	18.0	18.7	19.8	37.5	24.0	49.0	4.2	15.6	31.2
Boar	25.3	26.5	30.7	17.5	19.8	17.6	39.6	23.0	47.3	1.0	18.7	33.0
Elk	26.5	26.5	28.6	18.4	18.4	19.4	39.4	22.4	47.9	4.1	18.4	29.6
Eurasian lynx	26.5	25.5	34.0	14.0	20.25	19.25	34.0	26.5	39.4	7.4	8.5	44.7
Human	24.0	24.0	30.0	22.0	20.0	17.5	36.2	26.0	33.0	4.4	14.3	48.3
Quail	20.9	26.4	29.7	23.0	22.0	19.8	35.2	23.0	37.4	3.3	9.9	49.4
Python molurus	29.0	21.0	31.0	19.0	21.0	16.5	40.0	22.5	43.4	1.6	10.0	45.0
Frog	21.6	25.7	32.4	20.3	24.3	23.0	39.2	13.5	27.0	6.8	25.7	40.5
Standard deviation	1.95				1.29				4.83			
Bias	0.09				0.17				0.4			

of nucleotides in DNA sequence is in close relation to the extreme conservation of cytochrome *b* protein. The degeneracy of the genetic code regards mostly the third codon position. The nucleotide substitutions in this position are predominantly silent and thus are tolerated by natural selection. Changes at the first codon position more often results in amino-acid substitution, therefore this codon position is less variable. Moreover, it has been shown that cytochrome *b* gene sequence of some species has a unique composition differing from that in other species (Irvin *et al.*, 1991). In the study mentioned above dealing with 20 species of mammals, the human and zebra sequences were unique in preferring C over A in the light DNA strand. From our research it seems that this phenomenon is more common and it concerns not only mammals. We observed the C over A preference for the quail, frog, python and elk sequences (Table 2). Interestingly, these species represent

Table 2. Contents of adenine (A) and cytosine (C) in analysed fragment of cytochrome *b* gene (%, light strand).

Species	A	C
European bison	30.63	24.72
American bison	30.33	27.00
Polish Red cattle	31.02	24.82
Polish Whiteback cattle	31.02	24.82
Zebu	31.02	24.82
Goat	30.70	22.30
Sheep	31.49	24.22
Boar	30.91	24.36
Human	31.13	28.93
Eurasian lynx	30.51	24.41
Elk	25.55	32.12
Quail	28.03	28.72
Python molurus	26.79	31.79
Frog	24.22	25.11

distant systematic groups. Since there is little information about the base and amino-acid composition in amphibians, reptiles and birds, further investigations, including more representatives, should be carried out to determine if this phenomenon is typical or common for certain groups. The opposite bias (A over C) was observed in all other species considered. We observed that all the analysed species belonging to the *Artiodactyla* order have fewer pyrimidines in the light strand than the other species (Table 3). This is also in agreement with the results of Irvin *et al.*

Table 3. Contents of purines (A+G) and pyrimidines (T+C) in analysed fragment of cytochrome *b* gene (%, light strand).

Species	A+G	T+C
European bison	47.2	52.8
American bison	46.0	54.0
Polish Red cattle	47.4	52.6
Polish Whiteback cattle	47.4	52.6
Zebu	47.4	52.6
Goat	49.1	50.9
Sheep	48.4	51.6
Boar	45.8	54.2
Elk	44.4	55.6
Eurasian lynx	46.8	53.2
Human	41.2	58.8
Quail	45.3	54.7
Frog	43.2	56.8
Python molurus	42.6	57.4
Mean value for artiodactyls	47.02	52.98
Mean value for nonartiodactyls	43.82	56.18

(1991). Therefore, the question rises of how the observed preference in base composition is connected with the structure and function of cytochrome *b* protein. Translation products predicted from the cytochrome *b* gene sequences of all 14 species are of 91 amino acids in length. The overall rate of amino-acid identity is about 61% (not shown). Similar observations have been made by other authors for the complete cytochrome *b* gene (Irvin *et al.*, 1991). On the basis of mutational and evolutionary studies, it has been assumed that this protein contains both redox centers Q_o and Q_i involved in electron transfer (Hatefi, 1985; Howell & Gilbert, 1988). The fragment of cytochrome *b* gene under consideration in the present study codes for the first part of the second outer segment, the first and the second transmembrane segments and the second inner segment of cytochrome *b* protein (amino acids 50–143) (Fig. 2). Most of the outer segments are occupied by the Q_o redox

center. On the basis of the structural model of cytochrome *b* adapted from Brasseur (1988) (Fig. 2), we observe that the region covering the Q_o center is extremely conservative in the animal kingdom and there are no changes in preserving the metabolic energy supplies for living organisms. Additionally, the performed studies allowed for a comparison of the results obtained for the most often tested mammal species (e.g. human, rat, mouse, pig) with

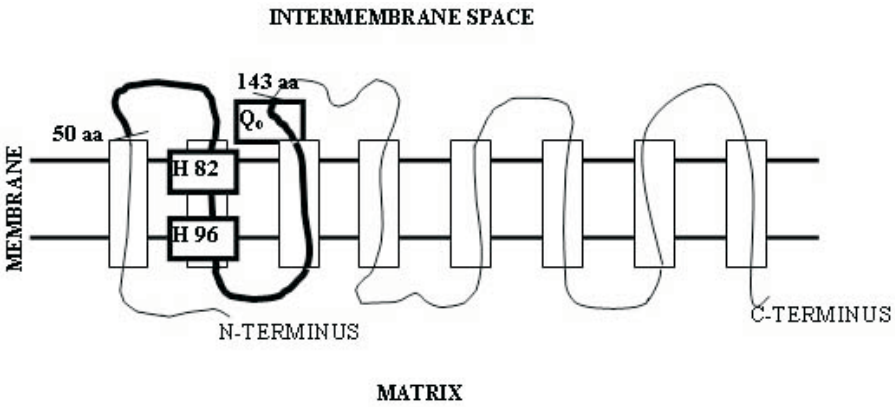

Figure 2. Structural model of cytochrome *b* (adapted from Brasseur, 1988).

Histidines H 82 and H 96 – the heme ligands; Q_o – part of Q_o redox center; bolded line – positions corresponding to amino acids coded by nucleotide sequence analysed at the present study.

the amino-acid sequences analysed in the present study (Fig. 1). This is connected with the high importance of the Q_o redox center in metabolic respiration and in turn, with a considerably slower evolution rate compared to other parts of the cytochrome *b* protein (Howell, 1989; Irvin *et al.,* 1991). Other positions of high evolutionary conservatism in cytochrome *b* are the four histidines which are the heme ligands. Histidines 82 and 183 are heme b_L ligands and histidines 96 and 197 are heme b_H ligands (L stands for low and H for high oxido-reduction-midpoint-potential) (Brasseur *et al.,* 1999). Two of them, *i.e.* His 82 and His 96, were covered by the region under consideration here. We observed that these amino-acid positions are absolutely conserved in all tested species. We observed only five nucleotide differences at the third position of the codon for His 82, and three differences at the third position for His 96 among the 14 species analysed. The performed studies revealed that there is a strong non-random distribution of nucleotides in the cytochrome *b* gene in all tested species. This is correlated with the biological function of the protein and evolution pattern directed on those obtained for typical Polish species and breeds. Regarding the level of interspecies variability in the analysed fragment of cytochrome *b* gene, the results showed that the determined sequences and methods applied can be also used in identification of variable biological samples. Thus we hope that the results of this study constitute a step in determining the genetic characteristics of valuable, rare Polish species.

REFERENCES

Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. (1981) Sequence and organization of the human mitochondrial genome. *Nature.*; **290**: 41457–65.

Bartlett SE, Davidson WS. (1992) FINS (forensically informative nucleotide sequencing): A procedure for identifying the animal origin of biological specimens. *BioTechniques.*; **12**: 408–11.

Brasseur G, Brivet-Chevillotte P, Lemesle-Meunier D, Di Rago JP. (1999) The

bc1 complex in the mitochondrial respiratory chain. In *Mitochondrial Diseases. Models and Methods*. Liestienne P. ed., pp 97–113. Springer-Verlag, Berlin-Heidelberg.

Brasseur R. (1988) Calculation of three-dimentional structure of *Saccharomyces cerevisiae* cytochrome *b* inserted in a lipid matrix. *J Biol Chem.*; **263**: 12571–5.

Brown WM. (1985) The mitochondrial genome of animals. In *Molecular Evolutionary Genetics*. MacIntyre RJ. ed., pp 95–130. Plenum Press, New York, London.

di Rago JP, Netter P, Slonimski PP. (1990) Pseudo-wild type revertants from inactive apocytochrome *b* mutants as a tool for the analysis of the structure/function relationships of the mitochondrial ubiquinol-cytochrome *c* reductase of *Saccharomyces cerevisiae*. *J Biol Chem.*; **265**: 3332–9.

Fukuda M, Wakasugi S, Tsuzuki T, Nomiyama H, Shimada K, Miyata T. (1985) Mitochondrial DNA-like sequences in the human nuclear genome. Characterization and implications in the evolution of mitochondrial DNA. *J Mol Biol.*; **186**: 257–66.

Hall TA. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.*; **41**: 95–8.

Hatefi Y. (1985) The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem.*; **54**: 1015–69.

Howell N, Gilbert K. (1988) Mutational analysis of the mouse mitochondrial cytochrome *b* gene. *J Mol Biol.*; **203**: 607–18.

Howell N. (1989) Evolutionary conservation of protein regions in the proton-motive cytochrome *b* and their possible roles in redox catalysis. *J Mol Evol.*; **29**, 157–69.

Irwin DM, Kocher TD, Wilson AC. (1991) Evolution of the cytochrome *b* gene of mammals. *J Mol Evol.*; **32**: 128–44.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primes. *Proc Natl Acad Sci USA.*; **86**: 6196–200.

Lestienne P, Bouzidi M, Desguerre I, Ponsot G. (1999) Molecular basis of mitochondrial diseases. In *Mitochondrial Diseases. Models and Methods*. Lestienne P. ed., pp 33–58. Springer-Verlag, Berlin-Heidelberg.

Li WH, Lou CC, Wu CI. (1985) Evolution of DNA sequences. In *Molecular Evolutionary Genetics*. MacIntyre RJ. ed., pp 1–94. Plenum Press, New York, London.

Montgelard C, Catzeflis FM, Douzery E. (1997) Phylogenetic relationships of artiodactyls and cetaceans as deduced from comparison of cytochrome *b* and 12S rRNA mitochondrial sequences. *Mol Biol Evol.*; **14**: 550–9.

Parson W, Pegoraro K, Niederstatter H, Foger M, Steinlechner M. (2000) Species identification by means of the cytochrome *b* gene. *Int J Legal Med.*; **114**: 23–8.

Prusak B, Grzybowski G, Zięba G. (2004) Taxonomic position of *Bison bison* (Linnaeus, 1758) and *Bison bonasus* (Linnaeus, 1758) based on analysis of *cytb* gene. *Anim Sc Pap Rep.*; **22**: 27–35.

Shoffner JM, Wallace DC. (1990) Oxidative phosphorylation diseases: disorders of 2 genomes. *Adv Hum Genet.*; **19**: 267–73.

Sueoka N. (1988) Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci USA.*; **85**: 2653–7.

Trumpower BL. (1990) Cytochrome-bc1 complexes of microorganisms. *Microbiol. Rev.*; **54**: 101–29.

Wilson MR., Polanskey D, Butler J, DiZinno JA, Repogle J, Budowle B. (1995) Extractoin, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *Bio Techniques.*; **18**: 662–9.

Zehner R, Zimmermann S, Mebs D. (1998) RFLP and sequence analysis of the cytochrome *b* gene of selected animals and man: methodology and forensic application. *Int J Legal Med.*; **111**: 323–7.