Ann. Anim. Sci., Vol. 18, No. 1 (2018) 3–15 *DOI: 10.1515/aoas-2017-0016*

DE GRUYTER
OPEN

A MOLECULAR CHARACTERISTIC OF THE ANATIDAE MITOCHONDRIAL CONTROL REGION – A REVIEW*

Joanna Warzecha♦, Agnieszka Fornal, Maria Oczkowicz, Monika Bugno-Poniewierska

Department of Animal Molecular Biology
National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland
♦Corresponding author: joanna.warzecha@izoo.krakow.pl

Abstract

Mitochondrial DNA (mtDNA) is a molecular tool that is very effective in genetic research, including phylogenetic analysis. The non-coding region is the most variable fragment of mtDNA, showing variability in length and nucleobase composition and containing three domains: two hypervariable peripheral regions and the conserved domain (D-loop) in the middle. The Anseriformes are amongst the best studied avian groups, including approximately 150 species and containing geese, swans, ducks (Anatidae), the Magpie goose (Anseranatidae) and screamers (Anhimidae). The most numerous family is the Anatidae, appearing in close relationships within the phylogenetic branches of the species. There are differences between the non-coding region of the Anatidae in comparison to other avian control regions. In the article presented below the control region sequences and the phylogeny of the Anatidae were reviewed.

Key words: mitochondrial DNA, control region, Anseriformes, Anatidae, goose, swan, duck

Phylogenetics and the molecular evolution of vertebrates have been widely investigated, particularly by using mitochondrial DNA (Ramirez et al., 1993). Despite its quick evolution, mitochondrial DNA is conservative in length, gene content and organisation. Additionally, it is a commonly used molecular tool in biological identification research, taxonomy and phylogeny in living organisms (Brown et al., 1982; Ramirez et al., 1993; Castro et al., 1998; Slack et al., 2003; Bucheli and Wenzel, 2005). Mitochondrial DNA encodes polypeptides of subunits of the enzyme complex of the oxidative phosphorylation – a main process of generating adenosine triphosphate (ATP), the energy source of the cell (Allen, 1996; Boore, 1999; Wallace,

*Acknowledgements: The work financed from the research activities of the National Research Institute of Animal Production, project no. 04-011.1.

2005). Mitochondrial DNA consists of approximately 15–20 kb in animals and it is a double-stranded circular genome (Boore, 1999). Heavy-strand (H), rich in purines, encodes most genetic information with genes for 12 polypeptides, 14 tRNA and 2 rRNA and the second, light-strand (L) that is rich in pyrimidine, codes only 1 polypeptide and 8 tRNA (Taanman, 1999; Holland and Parsons, 1999). There are between 2 and 10 copies of mtDNA in the single mitochondrion and their number depends on the individual cell function. For example, the number of mtDNA copies in oocytes is remarkably high ($\geq 10^5$), whereas this amount is much lower (50–75) in the sperm cells, so it has been recognised that the mitogenome is maternally inherited (Taanman, 1999; Solano et al., 2001).

Characteristics and comparison of the mammals and avian control region

Mitochondrial DNA has a conserved set of 13 genes for coding proteins: ND1-6, ND4L, ATP6, ATP8, COI-III, cytb, and 22 tRNA, 2 rRNA (12S rRNA and 16S rRNA). This arrangement of vertebrate genes is typical and their length is similar among species (Jiang et al., 2010; Liu et al., 2013). The genes are located closely, without intronic sequences in between, separated with short sequences of tRNA. mtDNA contains one non-coding region, known as the control region, and also known as the displacement loop (D-loop) in vertebrates because of the short three-stranded displacement loop structure created by a short nucleic acid strand displacing the H-strand complementary to the L-strand (Saccone et al., 1991; Taanman, 1999). The control region is the most variable fragment of mtDNA and the frequency of mutations may reach 5- to 10-fold more substitution when compared to other sequences in the mitogenome. It is also the most commonly used marker for studying the structures of the population and phylogeny of closely related species (Ruokonen and Kvist, 2002). It occurs in variable length, including the promoters for the transcription of the heavy-strand (HSP) and the light-strand (LSP) and the heavy-strand replication origin (O_H) (Saccone et al., 1991; Wolstenholme, 1992; Boore, 1999).

Most vertebrates have the control region located between tRNA-Phe and tRNA-Pro, whereas in most of the avian species it is bounded between tRNA-Glu and tRNA-Phe (Saccone et al., 1991; Ruokonen and Kvist, 2002) or between those for tRNAthr and tRNApro in some species of Picidae, Passeriformes, Falconiformes and Cuculidae (Mindell et al., 1998). The control region can be divided into three domains based on the position and frequencies of variable nucleotides (Brown et al., 1986; Saccone et al., 1991). Flanking domains are the most variable among mammals. Domain I includes sequences responsible for a termination of replication (ETAS – extended termination-associated sequence, including the shorter TAS sequences) of the H-strand and often includes VNTRs, while domain III contains LSP and HSP. The origin of the H-strand and three conserved sequence blocks occurs a high rate of nucleotides substitutions, insertions and deletions (indels) and VNTRs (Randi and Lucchini, 1998). The central domain II displays comparatively higher conservatism and is characterised by a high frequency of guanine and a lower one of adenine in the L-strand compared to the other two domains (Ruokonen and Kvist, 2002).

The composition and pattern of nucleotide in D-loop between mammalian and avian species are similar (Randi and Lucchini, 1998). In mammals, the D-loop shows high variability in length and base organisation (Saccone et al., 1991), whereas according to Ruokonen and Kvist (2002) the length of the investigated avian control region varies between 1028 bp (*Struthio camelus*) and 1581 bp (*Cyanoramphus auriceps*), with a wide range of diversity in genetic distances from 0.00 to 37.88% among species. The average nucleotide frequency in the avian non-coding region is A – 28.46%, C – 28.88%, G – 13.94%, and T – 28.71%, besides which the contents of cytosines and adenines are dominant in domain I and domain III, and cytosines and thymines in domain II. The greatest variability occurs in domain I in most genera of avian species (*Alectoris*, *Anser*, *Cepphus*, *Geospiza*, *Cyanoramphus*, *Grus*), but in many the high variable is domain III, or domain I and III are equal, respectively. The conserved sequences are concentrated in domain II, but also localised in all three domains in mammals and birds (Sbisà, 1997; Ruokonen and Kvist, 2002). In avian species the domain II length varies from 384 bp in the *Geospiza scandens* to 468 bp in the *Anser* and *Alectoris*, while in mammals it reaches a length between 304 and 328 bp.

Ruokonen and Kvist (2002) identified three conserved sequence boxes B, D and F in domain II; additionally, there was some similarity of the C box and E box, but only in some species, emphasising that the B box among the avian species is different to that in vertebrates and also that the D box and F box have better sequence matching compared to mammalian sequences. Conserved sequences outside of the central domain, CSB-1 (conserved sequence block 1) and cytosine stretch, are located in domain III and domain I, respectively. C-stretch is located at the beginning of domain I, creating a hairpin structure with the short region downstream, occurring in Galliformes and Anseriformes. Besides that, in domain I there were identified 16 of short sequence motifs – TAS – which are affected during the DNA synthesis termination and display the variable conservatism. CSB-1 in birds is the beginning of domain III, and differences occur in the sequence compared to mammals. CSB-2 and CSB-3 appear additionally in the mammalian control region, whereas in avian species these sequences have not been identified or are in an incomplete form (Figure 1) (Sbisà, 1997; Ruokonen and Kvist, 2002).

Analysis of the control region in the Anatidae

Anseriformes is a varied group of avian species displaying worldwide plurality and morphological and biological diversification (Olson and Feduccia, 1980; Gonzalez et al., 2009). The taxonomy of Anseriformes is complicated due to the close relations between about 150 species; nevertheless, it is one of the best studied avian groups (Donne-Goussé et al., 2002). The global population of Anseriformes is divided into the three families: Anhimidae (screamers), Anatidae (ducks, geese, swans), and Anseranatidae (magpie geese) (Donne-Goussé et al., 2002; Liu et al., 2013; IOC World Bird List). Anseriformes with Galliformes birds – quail (*Coturnix japonica*) and chicken (*Gallus gallus*) – form a sister clade – Galloanseres (Feinstein, 2006).

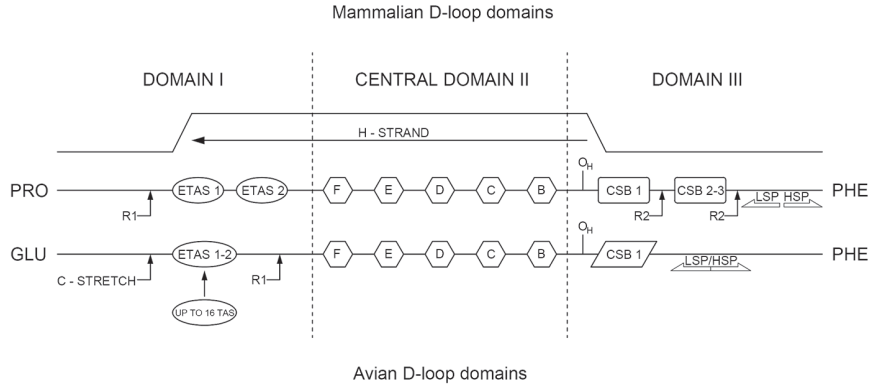

Figure 1. Schematic comparison of mitochondrial DNA non-coding region organization in mammals and avian. C-stretch – a stretch of cytosines, ETAS – extended termination-associated sequence domain and mammalian consensus sequence (including the shorter TAS sequences) (Sbisà, 1997); R1, R2 – tandem repeats in the peripheral domains; B, C, D, E, F – conserved sequence blocks of the central domain; O_H – origin of H-strand replication; CSB 1, -2 and -3 – conserved sequence blocks; LSP – light-strand transcription promoter; HSP – heavy-strand transcription promoter (according to Randi et al., 1998; Ruokonen and Kvist, 2002).

Their plurality and significance from an economic perspective show the Anatidae as an important part of Anseriformes. The Anatidae is separated into two subfamilies Anatinae (Mergini, Anatinae, Tadornini, Aythyini) and Anserinae (geese, swans, *Dendrocygna*). But there is a considerable debate about the classification status of some groups of Anseriformes and the phylogenetic relationships in some of the subfamilies, but the studies are still ongoing (Donne-Goussé et al., 2002; Gonzalez et al., 2009). Formerly, the relations among the families were created based on the phenotype of the animals – for example colour of their feathers, body form or size of the particular bird – but currently genetic analyses have replaced it. Currently, the analysis based on nuclear DNA and mitogenome clarifies the phylogenetic relationships among Anseriformes (Ruokonen et al., 2000). The control region of Anseriformes has a similar arrangement of genes and the attributes of a typical avian sequence. The presence of three domains has been reported by Donne-Goussé et al. (2002): domain I begins at the 5' end of the light-strand and contains 470 nucleotides, from position 471 begins the central domain to 1050, and finally domain III from position 1051 to the 3'end of non-coding region. The conserved central domain has three identifiable conserved sequence boxes – the F-, D- and C-box – and one conserved sequence of CSB-1 located in domain III similar to other birds.

The same as in other avian species, the control region begins with the sequence of tRNA-Glu and finishes with tRNA-Phe. The average size of the non-coding region of Anseriformes is in range of 1100 bp. Among Anatidae occurs a deletion of 100 – 130 bp versus other Anseriformes and also short deletions of about 1 – 20 bp in domain I compared to other species in these genera. Except for this, the deletions in the 5' (61 bp) and 3' (38 bp) end were reported in the Lesser snow goose and in the Peking duck compared to the domestic chicken. The highest average frequency of cytosine

(31.2%) was observed among Anseriformes D-loop, after adenine (28.0%), thymine (25.5%) and guanine (15.4%). A and C are the most common among domain I, C and T among conserved domain II, and A and T rich in domain II besides a very low amount of G, which is typical for avian species (Donne-Goussé et al., 2002).

Recent research has provided data on the evolution, genetic diversity and phylogeny of the Anatidae. The control region of Anatidae has a typical arrangement and order of genes beside the bird's usual nucleotide composition. It is placed in a characteristic location for most of avian, between tRNA-Glu and tRNA-Phe (Ruokonen and Kvist, 2002). The complete mitochondrial genome of the Swan goose *Anser cygnoides* is 16740 bp and the arrangement of the genes is typical for vertebrate animals containing 22 tRNA and 2 rRNA genes, 13 protein coding genes and one control region 1177 bp long, placed between tRNA-Glu and tRNA-Phe. The A (30.21%) and C (32.24%) are nucleotides with the highest frequencies (Zhu et al., 2015). The Yanling white goose *Anser cygnoides Linn. var domestica*, a famous goose breed of China, has mtDNA longer than Swan goose with the highest rates of A (30.22%) and C (32.02%) nucleotides as well. D-loop contains a typical structure and location for avian species and has 1178 bp length (Lin et al., 2014 a). The Xupu goose *Anser cygnoides Linn. var domestica,* indigenous domestic breed of geese, farmed in China for hundreds of years, has the same length of mitogenome as the Yanling white goose with the highest rates of A (30.21%) and C (32.02%). The composition of genes is characteristic for Anseriformes, containing one non-coding control region with a length of 1178 bp in a typical position (Lin et al., 2014 b). The complete mtDNA was sequenced in the case of the Wugangtong white goose *Anser cygnoides Linn. var domestica*, also a well-known breed in China. mtDNA of this goose diverges from the Swan goose with only one nucleotide base and has the highest content of A (30.22%) and C (32.02%) as well, including the same length of D-loop, located between tRNA-Glu and tRNA-Phe (Jiang et al., 2014). It was reported that the Bean goose *Anser fabalis* complete mitochondrial genome is shorter than previous mentioned breeds of geese, including the usual number of genes and one non-coding region (1182 bp) in typical location (Liu et al., 2013).

The Canada goose (*Branta canadensis*) is a common species in North America in which mtDNA sequences were used to specify genetic correlations among the various subspecies in North America. Approximately 80% of the control region is alignable when comparing *Anser* and *Branta* (Ruokonen et al., 2000). The mitogenome of the Canada goose *Branta canadensis* is 16760 bp and exhibits the typical gene arrangement of the Anatidae bird family. The control region is placed in the position 15567-16760 bp and it is located between tRNA-Glu and tRNA-Phe (Snyder et al., 2015). Moreover, in the publication of Donne-Goussé et al. (2002) and GenBank the sequences of the control region of all mentioned *Anser* and *Branta* species in this paper are deposited and listed in Table 1.

The Anatidae family, besides geese, also consists of ducks and swans. Although the mtDNA of many species of the family Anatidae has been sequenced, the phylogenetic relationships are incomplete. Closely related species reveal changes of mtDNA sequence, including the transitions in relation to transversions and also a higher proportion of silent substitutions (Ramirez et al., 1993). The ducks are more distantly

related to geese, the genetic distance between representatives of the Redhead duck *Aythya americana* and the White-fronted goose *Anser albifrons* is 15.7% (Feinstein, 2006). The mitochondrial DNA of the Whistling duck *Dendrocygna javanica* is 16753 bp in length and the GC content is slightly lower than other Anseriformes – the same as in the Magpie goose (Jiang et al., 2010). The gene order is similar to the Galloanserae birds and it is distinctive to avian species in the gene length of the H-strand and L-strand composition of the genes (Desjardins and Morais, 1990; Nishibori et al., 2001; 2002; 2004; Guan et al., 2009). The control region is 1187 bp and it is located between tRNA-Glu and tRNA-Phe, with typical base content, with the highest components of A (31.1%) and C (28.7%) nucleotides (Jiang et al., 2010). The sequences of the control region of the Fulvous whistling duck *Dendrocygna bicolor* is 1187 bp (AY112981) and the Eyton whistling duck *Dendrocygna eytoni* is 1183 bp long (AY112981) (Donne-Goussé et al., 2002; GenBank 2016). The species treated as endangered and classified on the IUCN Red List is the Baikal teal *Anas formosa*. The total mtDNA base composition is 16594 bp with a typical arrangement of genes and one non-coding region 1040 bp long. The overall base composition shows the highest content of A (29.52%) and G (25.76%) (Ryu and Hwang, 2011).

The Muscovy duck (*Carina moschata*) was domesticated in China and currently occurs around the world. In 2013, Tu and co-workers published a 16610 bp mtDNA sequence that included the 1040 bp D-loop between the t-RNA-Glu and tRNA-Phe rich in C (31.49%) and A (28.16%) nucleotides and disclosed a high (81.96%) similarity to the Mallard (*Anas platyrhychos*). Tu et al. also sequenced the 16604 bp mitochondrial DNA of the Xilin small partridge duck, the breed of duck also well-known in China (Tu et al., 2013). The gene arrangement is similar to the Anatidae and one non-coding region located in a quintessential position for the bird's sequences. The length of the control region is 1049 bp with a high content of A and T nucleotides (Xie et al., 2014).

The geese and the swans are related more closely to each other than to ducks; the genetic distance between the representatives of the goose *Anser albifrons* and the swan *Cygnus columbianus columbianus* reaches 11.6%. The complete sequence of mitogenome of the Whistling swan *Cygnus columbianus columbianus* is 16728 bp long, so that it is shorter than representatives of *Anser* species. The gene composition and many features are present also in *Gallus gallus* and is typical for birds – whereas the control region location is in the position 1891-3049 bp (Feinstein, 2006). The separate subspecies of *Cygnus columbianus columbianus* is *Cygnus columbianus bewickii,* known as the Bewick's swan. Despite the fact that these two species occur under the collective name of the Tundra swan, it is easy to distinguish them by their morphological traits. The complete mitochondrial DNA of the Bewick's swan is one nucleotide shorter than the Whistling swan, having the distinctive order of genes typical for birds and also identical to other representatives of *Cygnus* (Jiang et al., 2010; Lee et al., 2012). The length of the non-coding region is shorter than the Whistling swan (1157 bp) (Lee et al., 2012). On the other hand, the complete mtDNA sequence of *Cygnus columbianus jankowskii* is 16723 bp in length, with a dominance of A (30.0%) and C (31.9%) bases. The non-coding region is 1148 bp long and displays differences in length as compared to *Cygnus columbianus bewickii* (1157 bp) and

Cygnus columbianus columbianus (1159 bp), typically located between tRNA-Phe and tRNA-Glu for Anseriformes (Wang et al., 2014).

Table 1. The collation of mtDNA information of representatives of Anatidae family mentioned in the text

Species	Length of mtDNA [bp]	Length of D-loop [bp]	Content of nucleotides of mtDNA*			
			A	C	T	G
Swan goose	16740	1177	30.21%	32.24%	22.49%	15.06%
Yanling white goose	16742	1178	30.22%	32.02%	32.02%	15.70%
Xupu goose	16742	1178	30.21%	32.02%	22.70%	15.08%
Wugangtong white goose	16741	1177	30.22%	32.02%	22.69%	15.60%
Bean goose	16688	1182	30.06%	31.84%	22.74%	15.36%
Greylag goose	16738	1179	30.19%	32.14%	22.58%	15.09%
White-fronted goose	16737	1179	30.15%	32.05%	22.63%	15.18%
Lesser white-fronted goose	-	1172	-	-	-	-
Bar-headed goose	16731	1176	-	-	-	-
Ross's goose	-	1175	-	-	-	-
Emperor goose	-	1149	-	-	-	-
Canada goose	16760	1192	30.20%	32.10%	22.60%	15.10%
Red-breasted goose	-	1173	-	-	-	-
Barnacle goose	-	1183	-	-	-	-
Whistling duck	16753	1187	30.44%	30.44%	23.67%	15.45%
Fulvous whistling duck	-	1187	-	-	-	-
Eyton whistling duck	-	1183	-	-	-	-
Baikal teal	16594	1040	29.52%	15.52%	22.51%	25.76%
Muscovy duck	16610	1040	28.16%	31.49%	21.93%	16.11%
Xilin small partridge duck	16604	1049	29.20%	32.82%	22.19%	15.79%
Whistling swan	16728	1159	-	-	-	-
Bewick's swan	16727	1157	-	-	-	-
Cygnus columbianus jankowskii	16723	1148	30.00%	31.90%	22.80%	15.30%
Black swan	16748	1184	-	-	-	-
Mute swan	16739	1169	-	-	-	-

*Data not available marked by dash.

Other representatives of *Cygnus* genera are the Black swan *Cygnus atratus* and the Mute swan *Cygnus olor.* Their gene arrangement of mtDNA is similar to the Galloanserae and other vertebrate species. The mitochondrial DNA sequence of *Cygnus atratus* is 16748 bp in length and the G and C bases' are similar to Anseriformes; but it is higher than in Galliformes. The D-loop of *Cygnus atratus* is 1184 bp long and the compositions of the C (33.8%) and A (28.1%) bases are the most frequent (Jiang et al., 2010). The mtDNA of *Cygnus olor* has an overall length shorter than *Cygnus atratus*, with the bases content of A+T – 52.2% and G+C – 47.8% including the D-loop sequence 1169 bp long located between the tRNA-Glu and tRNA-Phe genes. After phylogenetic analysis, *Cygnus olor* is more closely related to *Anser cygnoides* than to any other species in the *Cygnus* genus (Table 1) (Park et al., 2015).

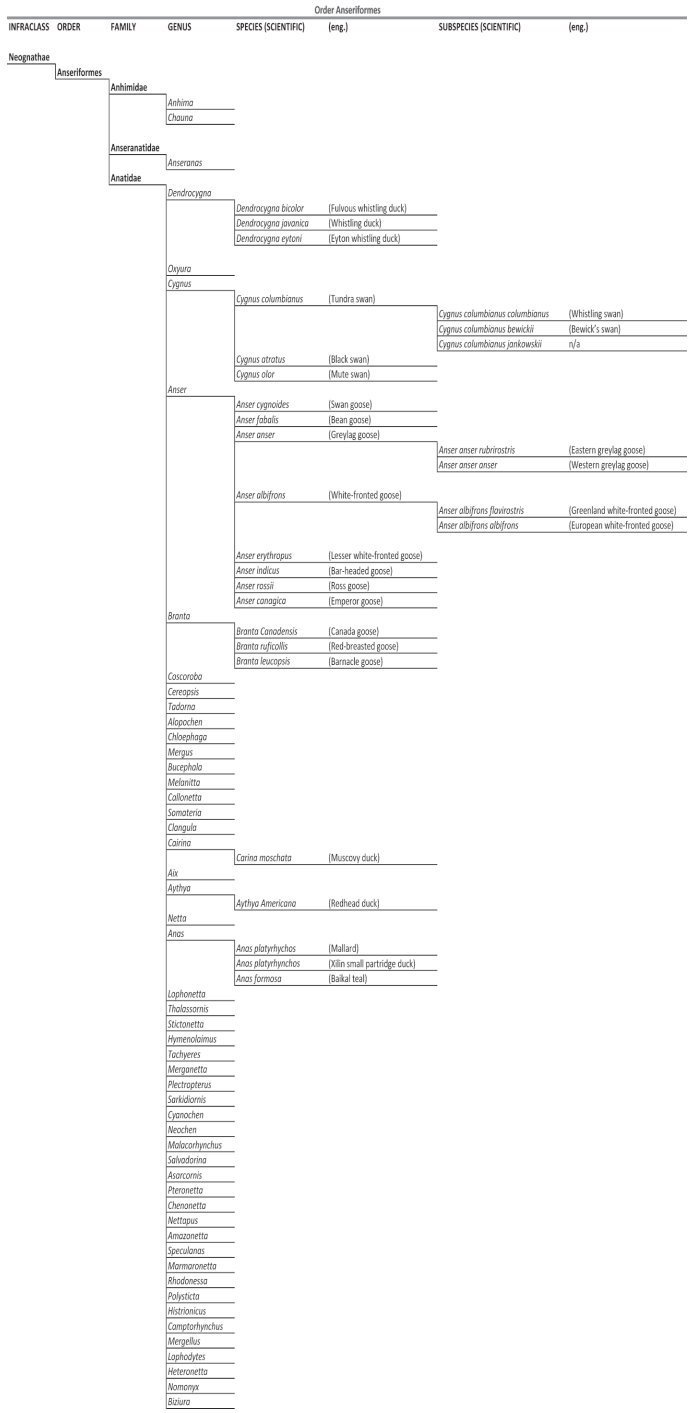

Figure 2. Phylogeny of the Anatidae family with species mentioned in this text (IOC World Bird List)

Phylogeny of the Anatidae family

Molecular analysis of the control region is a useful tool in constructing the relatively deep phylogeny in families such as Anseriformes, obtaining more stable, resolved and robust trees than the trees constructed using the protein coding genes of cytb or ND2. The conventional division in the most numerous group – the Anatidae – is between two subfamilies: the Anatinae and Anserinae, based on the analyses with high bootstrap values and indels mutations such as a large deletion in the control region of ca. 100 – 130 bp in the Anatinae compared to Anserinae (Liu et al., 2013; Donne-Goussé et al., 2002). Donne-Goussé et al. (2002) indicated five main clades within the Anatinae: (1) Anatini (*Anas*, *Lophonetta*), (2) Aythyini (*Aythya*, *Netta*), (3) Cairinini (*Cairina*, *Aix*) + Anatini + Aythyini + *Chenonetta* + *Marmaronetta*, (4) Margini (*Mergus*, *Bucephala*, *Melanitta*, *Callonetta*, *Somateria*, *Chloephaga*), and (5) Tadornini (*Tadorna*, *Alopochen, Chloephaga*) (Figure 2). The Anatini and Aythyini are linked sister tribes; the Cairinini form a monophyletic large clade with the Anatini, Aythyini, *Chenonetta* and *Marmaronetta*; the Margini is divided into six genera; and finally the Tadornini represent the first divergence among the Anatinae.

According to previous authors (Donne-Goussé et al., 2002; Livezey, 1997), the Anserine sub-family includes the Anserine (*Anser*, *Branta*) and Cygnini (*Cygnus*, *Coscoroba*, *Cereopsis*) tribes; among these tribes the Coscoroba and Cereopsis, the Coscoroba and Cereopsis are considered sister genera suggesting that Cygnus evolved earlier than the other genera. The *Anser* genus is clearly monophyletic and simultaneously closely related to the *Branta*. The genera *Dendrocygena* commonly referred to as whistling ducks diverged earlier than the Anatnae and Anserinae (Donne-Goussé et al., 2002; Livezey, 1997). Based on two mitochondrial genes *cytb* and *nd2*, Gonzalez et al. (2009) extended the phylogeny of Anseriformes suggesting that previous molecular phylogeny differences could be due to taxon sampling and that *Cereopsis* and *Coscoroba* evolved earlier than the genera *Anser*, *Branta* and *Cygnus* contrary to the results reported by Donne-Goussé et al. (2002).

The Oxyurini (Stiff-tailed ducks), including the genera *Oxyura*, *Nomonyx*, *Bizura* and *Heteronetta*, is a monophyletic group considered as a non-natural closely related group to the Anserinae but not to the Anatinae. Additionally, *Malacorhynchus membranaceus*, known as the Pink-eared duck, encompasses a basal position to swans and geese; moreover, the genera *Netta* and *Tadorna* were reported to be non-monophyletic and the relationship between the *Callonetta* and *Melanitta* was not obtained. Aves are prone to hybridisation producing new hybrid individuals in the species and as such close relations are occurring, especially among Anseriformes, e.g. *Anser* and *Anas,* but also in many distantly-related species as well (Gonzalez et al., 2009). According to Heikkinen et al. (2015), the Greylag and domestic goose hybridised, resulting in phenotypical hybrids.

The *Anser* and *Branta* form a tribe, Anserini, with a high grade of alignment reaching almost 80%. The branching order and grouping of the *Anser* is not clear because of the close relatedness among the species in general, whereas the genetic distances vary from 0.9% to 5.5%. In contrast, the distances between the subspecies of the European and Greenland white-fronted goose reach 0.8%, and for the Western

and Eastern greylag goose it is 1.0%. The most divergent taxa are the Snow goose and Ross's goose, creating a clade with high bootstrap support in the maximum parsimony and the maximum likelihood method, but at the same time these geese show the smallest divergence at a species level (0.3%). The Eastern and Western greylag goose are monophyly, with high support of both the aforementioned methods as well. In addition, relatively high bootstrap values are present in the monophyletic Pink-fronted and White-fronted geese as well as the maximum parsimony and the maximum likelihood method, however the Lesser white-fronted, Pink-footed, and Bean goose form a polytomy (Ruokonen et al., 2000). This close relationship between the White-fronted, Pink-footed and Bean goose obtained the same results as the monophyletic group of the Bar-headed, Greylag and Ross's goose (Lee et al., 2008). The Bean goose *Anser fablis* is closely related to the *Anser albifrons, Anser anser* and *Anser cygnoides* tribal members (Liu et al., 2013). Moreover *Anser cygnoides* is more closely related to *Anser anser* than to *Anser fabalis* and *Anser albifrons* (Liu et al., 2015).

Heikkinen et al. (2015) reported 44 haplotypes among wild Greylag and domestic geese population, of which six belong to domestic geese and 35 to wild Greylag geese, whereas three haplotypes were shared. Two main haplotypes were obtained for the domestic goose, D3 and D4, while haplotype diversity in the wild Greylag goose is very high (>0.8) in Eastern populations and in a similar range in European populations. For all Greylags, the diversity is 0.86 and for domestic geese it is 0.57, but for non-breed domestic geese, the diversity of haplotypes is higher than for most breeds. Based on six haplogroups, Heikkinen et al. (2015) reported that there were only minor differences in mitochondrial diversity between the graylag and European domestic geese. The close relatedness of the *Anser* species, as well as the short internal branches among the haplotypes of this species, indicate an existence of a period of rapid cladogenesis in the evolution of the *Anser* family. The speciation among *Anser* species dates back approximately 2–2.5 million years ago to the Late Pliocene and Pleistocene, whereas 0.5 million years ago, approximately in the Middle Pleistocene, marked the diversity of the most closely related *Anser* species (Ruokonen et al., 2000).

Conclusions

Domestic breeds occupy almost all biogeographical regions of the world and represent one of the major groups of birds, especially from an economic perspective, showing them to be a perfect phylogenetic and evolutionary model as well. Anseriformes are generally a monophyletic group forming a sister clade with Galliformes, but at the species level the close relationships between these domestic breeds remain complex and unsolved. Phylogenomics is useful for resolving and verifying such relationships, including mitochondrial genomes. The control region – as the most variable sequence of mtDNA – is of special interest in this type of research, as well as a valuable information tool for determining identification in different biological material such as plumage, tissues and fossils. The most numerous family is the Anatidae and many new reports about the phylogenetic of these birds are published each year. At present, most of the research reported out of Asia and the USA analyses

the Anatidae, primarily because many of these species are indigenous breeds and under federal protection. The D-loop of the Anatidae family is characterised for avian species by typical genes order in the location between tRNA-Glu and tRNA-Phe and base composition. The sequences of the control region among the *Anser*, *Cygnini*, *Dendrocygna* and *Anas* species range in size from 1049 to 1192 bp, indicating high diversity between the species. However, the phylogeny is still lacking so the research should be still ongoing for creating a relatively deep phylogeny and for providing further information into the diversity and evolution among this large and commonly bred group of poultry.

References

Allen J.F. (1996). Separate sexes and the mitochondrial theory of ageing. J. Theor. Biol., 180: 135–140.

Boore J.L. (1999). Animal mitochondrial genomes. Nucleic Acids Res., 27: 1767–1780.

Brown G.G., Gadaleta G., Pepe G., Saccone C., Sbisa 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., Prager E.M., Wang A., Wilson A.C. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol., 18: 225–239.

Bucheli S.R., Wenzel J. (2005). Gelechioidea (Insecta: Lepidoptera) systematics: a reexamination using combined morphology and mitochondrial DNA data. Mol. Phylogenet. Evol., 35: 380–394.

Castro J.A., Picornell A., Ramon M. (1998). Mitochondrial DNA: a tool for populational genetics studies. Int. Microbiol., 1: 327–332.

Desjardins P., Morais R. (1990). Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J. Mol. Biol., 212: 599–634.

Donne-Gousse C., Laudet V., Hänni C. (2002). A molecular phylogeny of anseriformes based on mitochondrial DNA analysis. Mol. Phylogenet. Evol., 23: 339–356.

Feinstein J. (2006). The mitochondrial genome of *Cygnus columbianus*, the Whistling Swan: Full Length Research Paper. DNA Sequence, 17: 99–106.

GenBank (www.ncbi.nlm.nih.gov/genbank), 22.10.2016.

Gonzalez J., Düttmann H., Wink M. (2009). Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. J. Zool., 279: 310–318.

Guan X., Silva P., Gyenai K.B., Xu J., Geng T., Tu Z., Samuels D.C., Smith E.J. (2009). The mitochondrial genome sequence and molecular phylogeny of the turkey, *Meleagris gallopavo*. Anim. Genet., 40: 134–141.

Heikkinen M.E., Ruokonen M., Alexander M., Aspi J., Pyhäjärvi T., Searle J.B. (2015). Relationship between wild greylag and European domestic geese based on mitochondrial DNA. Anim. Genet., 46: 485–497.

Holland M.M., Parsons T.J. (1999). Mitochondrial DNA sequence analysis – validation and use for forensic casework. Forensic Sci. Rev., 11: 21–50.

IOC World Bird List (www.worldbirdnames.org/ bow /waterfowl) 30.03.2016.

Jiang F., Miao Y., Liang W., Ye H., Liu H., Liu B. (2010). The complete mitochondrial genomes of the whistling duck (*Dendrocygna javanica*) and black swan (*Cygnus atratus*): dating evolutionary divergence in Galloanserae. Mol. Biol. Rep., 6: 3001–3015.

Jiang G.T., Lin Q., He P., Zhang X., Yun L., Li X., Liu G., Li C., He X., Dai Q.Z. (2014). The complete mitochondrial genome of the Wugangtong white goose. Mitochondr. DNA, 1–3: 3890–3892.

Lee J.H., Ryu S.H., Kang S.G., Hwang U.W. (2012). Complete mitochondrial genome of the Bewick's swan *Cygnus columbianus bewickii* (Aves, Anseriformes, Anatidae). Mitochondr. DNA, 23: 129–130.

Lee S.Y., Scott G.R., Milsom W.K. (2008). Have wing morphology or flight kinematics evolved for extreme high altitude migration in the bar-headed goose? Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology, 148: 324–331.

Lin Q., Cao R., Jiang G.T., Liu G., Dai Q.Z., Zhang S.R., Hou D.X., He X. (2014 a). Determination and analysis of the complete mitochondrial genome sequence of Yanling white goose. Mitochondr. DNA Part A, 27: 1012–1013.

Lin Q., Cao R., Jiang G.T., Liu G., Qiu L., Hu G.B., Dai Q.Z., Zhang S.R., Hou D.X., He X. (2014 b). The complete mitochondrial genome sequence of the Xupu goose. Mitochondr. DNA Part A, 27: 1010–1011.

Liu C.Z., Wei G.H., Hu J.H., Liu X.Y. (2015). Complete mitochondrial genome of the Swan Goose (*Anser cygnoides L.*) and its phylogenetic analysis. Mitochondr. DNA Part A, 27: 2427––2428.

Liu G., Zhou L., Zhang L., Luo Z., Xu W. (2013). The complete mitochondrial genome of Bean goose (*Anser fabalis*) and implications for Anseriformes taxonomy. PLOS ONE, 8: e63334.

Livezey B.C. (1997). A phylogenetic analysis of basal Anseriformes, the fossil Presbyornis, and the interordinal relationships of waterfowl. Zool. J. Linn. Soc-Lond., 121: 361–428.

Mindell D.P., Sorenson M.D., Dimcheff D.E. (1998). Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci, 95: 10693–10697.

Nishibori M., Hayashi T., Tsudzuki M., Yamamoto Y., Yasue H. (2001). Complete sequence of the Japanese quail (*Coturnix japonica*) mitochondrial genome and its genetic relationship with related species. Anim. Genet., 32: 380–385.

Nishibori M., Tsudzuki M., Hayashi T., Yamamoto Y., Yasue H. (2002). Complete nucleotide sequence of the *Coturnix chinensis* (Blue-breasted quail) mitochondorial genome and a phylogenetic analysis with related species. J. Hered., 93: 439–444.

Nishibori M., Hayashi T., Yasue H. (2004). Complete nucleotide sequence of *Nurrada meleagris* (Helmeted guineafowl) mitochondrial genome. J. Poultry Sci., 41: 259–268.

Olson S.L., Feduccia A. (1980). Presbyornis and the origin of the Anseriformes (Aves: Charadriomorphae). Smithsonian Institution Press, 323.

Park C.E., Park G.S., Kwak Y., Hong S.J., Khan A.R., Jung B.K., Park Y.J., Kim J.G., Park H.C., Shin J.H. (2015). Complete mitochondrial genome of *Cygnus olor* (Aves, Anseriformes, Anatidae). Mitochondr. DNA Part A, 27: 3442–3443.

Ramirez V., Savoie P., Morais R. (1993). Molecular characterization and evolution of a duck mitochondrial genome. J. Mol. Evol., 37: 296–310.

Randi E., Lucchini V. (1998). Organization and evolution of the mitochondrial DNA control region in the avian genus Alectoris. J. Mol. Evol., 47: 449–462.

Ruokonen M., Kvist L. (2002). Structure and evolution of the avian mitochondrial control region. Mol. Phylogenet. Evol., 23: 422–432.

Ruokonen M., Kvist L., Lumme J. (2000). Close relatedness between mitochondrial DNA from seven Anser goose species. J. Evolution. Biol., 13: 532–540.

Ryu S.H., Hwang U.W. (2011). Complete mitochondrial genome of the Baikal teal *Anas formosa* (Aves, Anseriformes, Anatidae). Mitochondr. DNA, 22: 74–76.

Saccone C., Pesole G., Sbisá E. (1991). The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol., 33: 83–91.

Sbisà E., Tanzariello F., Reyes A., Pesole G., Saccone C. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene, 205: 125–140.

Slack K.E., Janke A., Penny D., Arnason U. (2003). Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene, 302: 43–52.

Snyder J.C., Mackaness C.A., Sopher M.R., Huber J.P., Disantis E.J., Senecal A.J., Byron P. Vaughn B.P., Desantis R.S., Tobelmann P.E., Hecht Balauff N.M., Barry P.M., Show M.D., Speering L.H., Genareo C.A., Brenner F.J., Durwood B.R. (2013). The complete mitochondrial genome sequence of the Canada goose (*Branta canadensis*). Mitochondr. DNA, 26: 672–673.

Solano A., Playán A., López-Pérez M.J., Montoya J. (2001). Genetic diseases of human mitochondrial DNA. Salud Pública Mex., 43: 151–161.

Taanman J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. BBA-Bioenergetics, 1410: 103–123.

Tu J., Si F., Wu Q., Cong B., Xing X., Yang F.H. (2013). The complete mitochondrial genome of the Muscovy duck, *Cairina moschata* (Anseriformes, Anatidae, Cairina). Mitochondr. DNA, 25: 102–103.

Wallace D.C. (2005). The mitochondrial genome in human adaptive radiation and disease: on the road to therapeutics and performance enhancement. Gene, 354: 169–180.

Wang J., Liu G., Zhou L., Qing H., Li L., Li B., Zhang L. (2014). Complete mitochondrial genome of Tundra swan *Cygnus columbianus jankowskii* (Anseriformes: Anatidae). Mitochondr. DNA Part A, 27: 90–91.

Wolstenholme D.R. (1992). Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol., 141: 173–216.

Xie Z., Zhang Y., Xie L., Deng X., Xie Z., Huang L., Huang J., Zeng T. (2014). Sequence and gene organization of the Xilin small partridge duck (Anseriformes, Anatidae, Anas) mitochondrial DNA. Mitochondr. DNA Part A, 27: 1579–1580.

Zhu H., Li B., Li L., Zhou L. (2015). Complete mitochondrial genome of Swan goose *Anser cygnoides* (Anseriformes: Anatidae). Mitochondr. DNA Part A, 27: 3177–3178.

Received: 15 XII 2016
Accepted: 20 VI 2017