Gene 534 (2014) 113–118

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

The complete sequence of the mitochondrial genome of the African Penguin (*Spheniscus demersus*)

Christiaan Labuschagne [a,c,*], Antoinette Kotzé [a,b,1], J. Paul Grobler [a,2], Desiré L. Dalton [a,b,1]

[a] Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa
[b] National Zoological Gardens of South Africa, P.O. Box 754, Pretoria 0001, South Africa
[c] Inqaba Biotechnical Industries (Pty) Ltd, P.O. Box 14356, Hatfield 0028, South Africa

ARTICLE INFO

Article history:
Accepted 16 September 2013
Available online 22 October 2013

Keywords:
Little blue penguin
Eudyptula minor
Rockhopper penguin
Eudyptes chrysocome
Next generation sequencing

ABSTRACT

The complete mitochondrial genome of the African Penguin (*Spheniscus demersus*) was sequenced. The molecule was sequenced via next generation sequencing and primer walking. The size of the genome is 17,346 bp in length. Comparison with the mitochondrial DNA of two other penguin genomes that have so far been reported was conducted namely; Little blue penguin (*Eudyptula minor*) and the Rockhopper penguin (*Eudyptes chrysocome*). This analysis made it possible to identify common penguin mitochondrial DNA characteristics. The *S. demersus* mtDNA genome is very similar, both in composition and length to both the *E. chrysocome* and *E. minor* genomes. The gene content of the African penguin mitochondrial genome is typical of vertebrates and all three penguin species have the standard gene order originally identified in the chicken. The control region for *S. demersus* is located between tRNA-Glu and tRNA-Phe and all three species of penguins contain two sets of similar repeats with varying copy numbers towards the 3′ end of the control region, accounting for the size variance. This is the first report of the complete nucleotide sequence for the mitochondrial genome of the African penguin, *S. demersus*. These results can be subsequently used to provide information for penguin phylogenetic studies and insights into the evolution of genomes.

1. Introduction

Mitochondrial DNA (mtDNA) is generally a 15–23 kb double-strand circular genome in animals and plays an important role in the process of metabolism and programmed cell death (Cao et al., 2006). This genome generally contains 13 protein-coding genes, two ribosomal RNAs, 22 transfer RNAs and a non-coding control region (D-loop) of variable length that contains signals required for replication and transcription (Ruokonen and Kvist, 2002; Wolstenholme, 1992). Being separate from the nucleus, the mitochondrial genome has several characteristics that make it unique; including maternal inheritance, its small size, fast evolutionary rate, limited recombination and relatively conserved gene content and organization (Avise, 1994; Brown, 1983; Cao et al., 2006; Wu et al., 2003). Due to these traits mtDNA has been used extensively for testing hypotheses of microevolution, studying population structure, phylogeography and phylogenetic relationships at various taxonomic levels (Cao et al., 2006; Zhou et al., 2009).

Since useful information can be identified from many of the mitochondrial genes and due to primers being functional for a wide range of taxa, the number of complete mitochondrial genomes is steadily increasing (Sammler et al., 2011). Complete mitochondrial genomes provide sets of genome-level characteristics, which are useful for modeling genome evolution and phylogenetic inference (Gibb et al., 2007; Lei et al., 2010). These characteristics include base composition, genetic codon variation, gene content and gene arrangement, tRNA and rRNA gene secondary structures and modes of replication and transcription (Lei et al., 2010). To date, complete mitochondrial genomes have been reported for only two penguin species, the little blue penguin (*Eudyptula minor*) and the Rockhopper penguin (*Eudyptes chrysocome*) (Slack et al., 2003; Watanabe et al., 2006). This study reports the complete mitochondrial genome of the African penguin (*Spheniscus demersus*) along with a comparative analysis of the complete mtDNA genome with the two other penguin species.

Abbreviations: A, adenosine; aa, amino acid; Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; ATP, adenosine 5-triphosphate; bp, base pair(s); C, cytosine; COX, cytochrome c oxidase; CR, control region; CSB, conserved sequence blocs; Cys, cysteine; Cytb, cytochrome b; DNA, deoxyribonucleic acid; G, guanine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; H, heavy; His, histidine; Ile, isoleucine; kb, kilobase; L, light; Leu, leucine; Met, methionine; min, minute; mtDNA, mitochondrial DNA; NAD, nicotinamide–adenine dinucleotide; NADH, nicotinamide–adenine dinucleotide (reduced); ND, NADH dehydrogenase; ng, nano gram; PCG, protein coding gene; PCR, polymerase chain reaction; Phe, phenylalanine; pM, pico mole; Pro, proline; RNA, ribonucleic acid; rRNA, ribosomal ribonucleic acid; s, second; S, subunit; Ser, serine; T, thymine; Thr, threonine; Tm, melting temperature; tRNA, transfer ribonucleic acid; Trp, tryptophan; Tyr, tyrosine; Val, valine; U, uracil.

* Corresponding author at: Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa. Tel.: +27 12 343 5829; fax: +27 12 343 0287.
E-mail addresses: Christiaan.Labuschagne@inqababiotec.co.za (C. Labuschagne), antoinette@nzg.ac.za (A. Kotzé), groblerjp@ufs.ac.za (J.P. Grobler), desire@nzg.ac.za (D.L. Dalton).
[1] Tel.: +27 12 339 2795.
[2] Tel.: +27 51 4013844.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.09.057

2. Methods and materials

2.1. DNA sample

A DNA sample prepared for a previous study on microsatellite development in *S. demersus* (Labuschagne et al., 2013) was used. The blood sample was from a captive breeding adult African penguin in a colony located in the KwaZulu-Natal Province of South Africa. Total genomic DNA was isolated using the Qiagen DNeasy® Blood and Tissue Kit.

2.2. Primer design, PCR amplification and DNA sequencing

Initial primers were designed based on 60 reads identified as mitochondrial from a next generation sequencing dataset (7706 reads) generated in a previous study (Labuschagne et al., 2013) on the GS FLX (Roche). The majority of the reads mapped to NADH dehydrogenase subunit 2 (38 reads) followed by NADH dehydrogenase subunit 6 (7 reads) and NADH dehydrogenase subunit 4 (6 reads). Gaps were then filled using the primer walking method. Assemblies, mapping and primer design were performed in CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark). Primer design parameters were set to a minimum melting temperature (Tm) of 53 °C, maximum Tm of 62 °C, primer length 16–21 bp and remaining settings on default. Assembly setting was set to auto-trim, minimum aligned read length of 30 bp, alignment stringency medium, ambiguity nucleotides and all other settings as default. All PCRs were performed utilizing an ABI 9700 thermal cycler (Applied Biosystems, Foster City, CA). Amplification reactions were done in a final volume of 25 μl containing 30 ng DNA, 25 pM of each primer and 2× DreamTaq® Green Master Mix (Thermo Fisher Scientific, Vilnius, Lithuania). Thermal cycling consisted of initial denaturation at 95 °C for 5 min, 45 cycles of denaturation at 95 °C for 30 s, annealing at 55–59 °C for 30 s, extension at 72 °C for 5 min, followed by final extension at 72 °C for 10 min. Resulting amplicons were inspected on 1% agarose gels followed by purification utilizing the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Orange, CA, USA). Purified templates were sequenced by utilizing a Big Dye V3.1 Terminator Kit (Applied Biosystems, Foster City, CA) and the ABI 3500XL genetic analyser (Applied Biosystems, Foster City, CA) following manufacturer's instructions.

2.3. Sequence assembly and sequence analysis

Sequences were checked, assembled and annotated in CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark). The boundaries of the protein-coding genes and rRNA genes were inferred by comparisons with the amino acid sequence of proteins and the nucleotide sequence of other birds including Flamingo (*Phoenicopterus roseus*; EF532932), Pacific loon (*Gavia pacifica*; AP009190), White stork (*Ciconia ciconia*; AB026818), Red-throated loon (*Gavia stellate*; AY293618), Little blue penguin (*E. minor;* AF362763) and Rockhopper penguin (*E. chrysocome*; NC_008138). The tRNA genes were identified by their cloverleaf secondary structure using tRNA-scan SE 1.21 (Lowe and Eddy, 1997) as well as ARWEN (online version) (Laslett and Canbäck, 2008) and verified by comparison with homologous sequences of other birds (mentioned above). Comparisons were made by forming assemblies between homologous sequences in CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark) using min aligned read length of 20 bp, alignment stringency low, ambiguity nucleotides and all other settings as default. The complete mtDNA sequence of *S. demersus* reported in this article was deposited in GenBank under accession number KC914350. CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark) was used to draw a maximum likelihood phylogeny between AB026818, AF362763, NC_008138 and KC914350 utilizing Neighbor Joining as starting tree algorithm, General Time Reversible as substitution model and bootstrapping of 1000 replicates.

3. Results and discussion

3.1. Mitochondrial genome organization

The complete mitochondrial genome of *S. demersus* as determined in this study, is 17,346 bp in length (Fig. 1), which is comparable to *E. chrysocome* (16,930 bp) and *E. minor* (17,611 bp). This length is not absolute, however, due to heteroplasmy caused by differences in the number of repeated motifs, ACAACAAACAACAA, at the 3′ end of the control region (CR). Heteroplasmy has also been reported in *E. minor* and *E. chrysocome* (Slack et al., 2003; Watanabe et al., 2006). *S. demersus* mtDNA genome shows 88.94% (91.77% excluding CR) similarity to *E. chrysocome* and 89.66% (91.75% excluding CR) similarity with *E. minor*, while *E. minor* and *E. chrysocome* have 87.26% (91.25% excluding CR) similarity. *S. demersus* and *E. minor* share a more recent common ancestor and group together but are both partitioned on a separate branch from *E. chrysocome*, as illustrated in Fig. 2. This observation is in agreement with our current understanding of their relationships. Baker et al. (2006) indicated that based on 2802 bp of nuclear and 2889 bp of mtDNA; *S. demersus*, *E. minor* and *E. chrysocome* diverged from the older Antarctic genera approximately 34–25 mya. The authors further indicated that *S. demersus*, *E. minor* grouped together, but were both partitioned from *E. chrysocome.* The gene content of the African penguin mt genome is typical of vertebrates, consisting of 13 protein coding genes (PCGs), 22 tRNAs and two rRNAs. As seen in other birds, NADH dehydrogenase subunit 6 and 8 tRNAs are transcribed from the light strand, while the other 12 protein coding genes, 14 tRNAs and two rRNAs are located on the heavy strand. Although variation in gene order has been described among avian mt genomes, all three penguin species have the standard gene order originally identified in the chicken (Desjardins and Morais, 1990). The nucleotide composition of the *S. demersus* mt genome (H strand) (A = 30.77%; C = 32.56%; G = 13.58%; T = 23.08%) is similar to that of *E. chrysocome* (A = 30.67%; C = 32.88%; G = 13.85%; T = 22.60%), *E. minor* (A = 30.96%; C = 31.84%; G = 13.53%; T = 23.67%) and other avian species. The A + T content of 53.85% is within range for avian mt genomes (51.6%–55.7%) and very similar to the other two penguin genomes (*E. chrysocome* = 53.27%; *E. minor* = 54.63%). One extra cytosine is present in NAD3 in all three penguin species. The extra nucleotide has been described in several other bird species as well as some turtles and is thought not to be translated (Mindell et al., 1998a,b). Russell and Beckenbach (2008) suggested that certain mitochondrial translation systems have the ability to tolerate frameshift insertions using programmed translational frameshifting, but the function of the extra nucleotide in NAD3 and its phylogenetic implications are still unclear (Kan et al., 2010).

3.2. Codon usage and sequence features of protein-coding genes

The usage of initial and termination signals as well gene length in comparison with two other penguin species is given in Table 1. The most common start codon is ATG. In COX1, all three penguin species as with most other birds (Slack et al., 2003) use the nonstandard start codon GTG. *E. minor* uses the same start codon for ND5, whereas the other two species use the standard ATG. The use of GTG in ND5 has also been described in ducks (Readhead duck, *Aythya americana*) and goose (Greater White-fronted Goose, *Anser albifrons*) (Slack et al., 2003). Furthermore, ATC (*S. demersus* and *E. minor)* and ATT (*E. chrysocome*) are used as start codons in ND3. This unusual start codon (isoleucine) has thus far only been found in ND3 in passerines (Watanabe et al., 2006). Stop codon usage is consistent for all three penguins across all 13 PCGs. As in the mtDNA genome of other birds, TAA is the most frequent stop codon. TAG is used for ND2, ND4 and ND6, while AGG was used for ND1 and COX1. Among neognath birds,

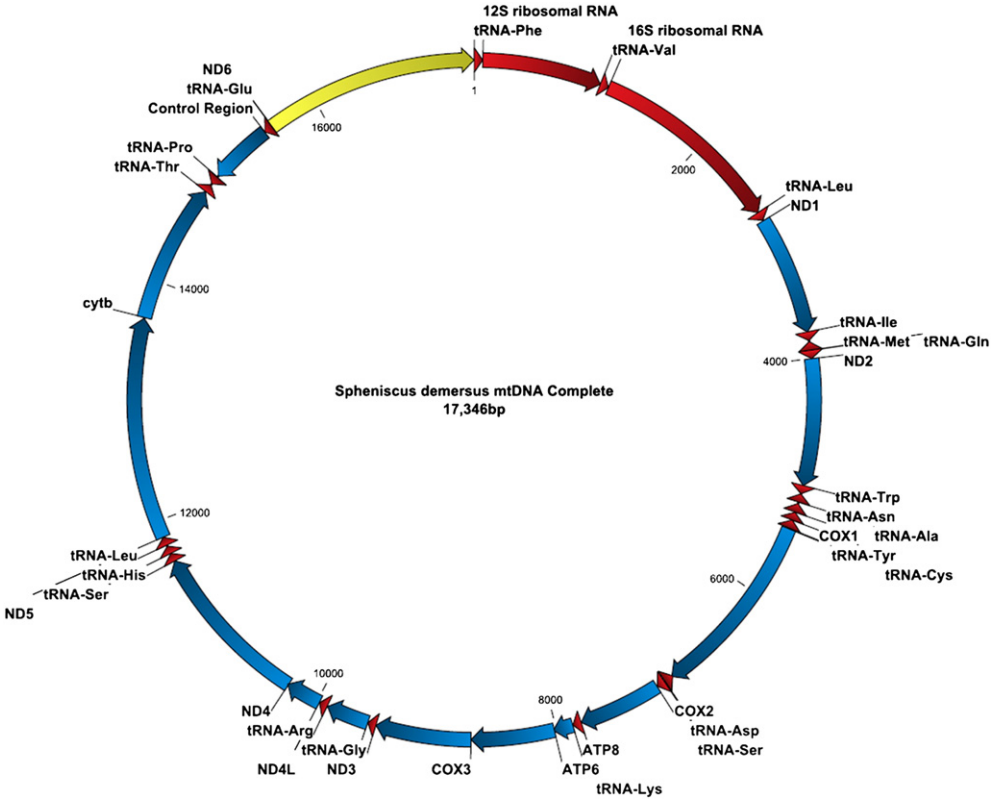

Fig. 1. Genetic map of the *S. demersus* mitochondrial genome. Annotation arrows indicate orientation of genes; COX1-3 indicates cytochrome oxidase subunits 1–3; ATP6/8, ATPase subunits 6–8; ND1-6/4L, NADH dehydrogenase 1-6/4L and CytB, cytochrome b. For designation of transfer RNAs, tRNAs and the three-letter code for amino acids is used.

ND4 is usually terminated through TAA or incomplete stop codons, TA- and T– (Slack et al., 2003). All three penguin species use the incomplete stop codon T– in COX3 as described in other birds. The terminal T serves as the stop signal after it is completed to UAA by post-transcriptional polyadenylation (Ojala et al., 1981). Identical gene length was observed among the three penguin species for 11 PCGs. *E. minor* contains one extra amino acid (aa) in ND5 (606 aa), while *E. chrysocome* contains one extra aa in Cytb (381 aa) when compared to the other two species. Varying sizes for ND5 have been reported previously with sizes ranging from 603 aa for the tinamou to 607 aa for a duck (Slack et al., 2003). Cytb size reports for birds are mostly 379 aa and 380 aa (Kan et al., 2010; Slack et al., 2003; Watanabe et al., 2006). The longest mtDNA PCG for all three species is NAD5, while the shortest is ATP8, as described in other birds. All three penguin species have one less aa than most other birds in both ATP8 and NAD6 (Slack et al., 2003).

3.3. Spacers and overlaps

A total of 19 intergenic spacers ranging from 1 bp to 1758 bp, are found in the mtDNA genome of *S. demersus* (Table 2). Among these, the longest non-coding region (1758 bp) is found between tRNA-Glu and tRNA-Phe and this will be discussed further under the control region section. Nineteen intergenic spacers are also found in *E. minor, while E. chrysocome* had only 18. Excluding the CR, the intergenic spacers amount to 80 bp in *S. demersus*, 64 bp in *E. chrysocome* and 60 bp in *E. minor*. The *S. demersus* mtDNA genome seems less compact when compared to the other two penguin species. In general, although length may vary, spacer and overlap positions are mostly conserved across the three penguin species. However, *S. demersus* contains a 8 bp spacer instead of an overlap observed in the other two species between tRNA-Ser(AGY) and tRNA-Leu(CUN). Furthermore, *E. minor* contains an

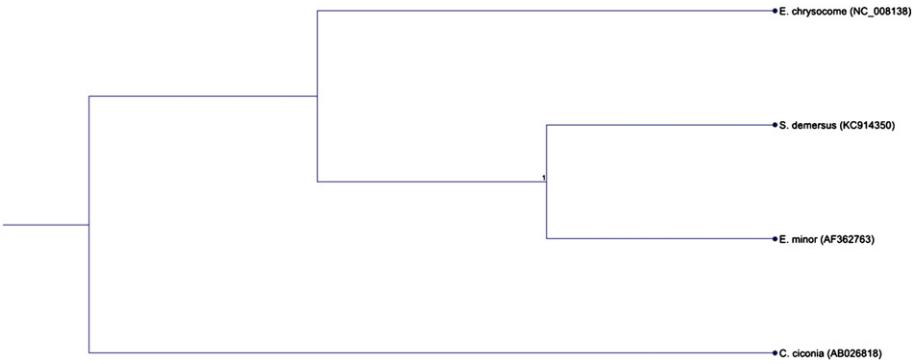

Fig. 2. Maximum Likelihood Phylogeny based on the mitochondrial genomes, excluding the control region showing the relationships among three penguin taxa and the white stork outgroup.

Table 1
Length and start/stop codons of mitochondrial protein-encoding genes of three penguin species.

Gene	Species	*S. demersus* (current study)	*E. chrysocome* Watanabe et al. (2006)	*E. minor* Slack et al. (2003)
ND1	Length (bases/amino acid)	978/325	978/325	978/325
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	AGG	AGG	AGG
ND2	Length (bases/amino acid)	1041/346	1041/346	1041/346
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAG	TAG	TAG
COX1	Length (bases/amino acid)	1551/516	1551/516	1551/516
	Start codon	GTG(Val)	GTG(Val)	GTG(Val)
	Stop codon	AGG	AGG	AGG
COX2	Length (bases/amino acid)	684/227	684/227	684/227
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAA	TAA	TAA
ATP8	Length (bases/amino acid)	165/54	165/54	165/54
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAA	TAA	TAA
ATP6	Length (bases/amino acid)	684/227	684/227	684/227
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAA	TAA	TAA
COX3	Length (bases/amino acid)	784/261	784/261	784/261
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	T–	T–	T–
ND3	Length (bases/amino acid)	352/116	352/116	352/116
	Start codon	ATC(Ile)	ATT(Ile)	ATC(Ile)
	Stop codon	TAA	TAA	TAA
ND4L	Length (bases/amino acid)	297/98	297/98	297/98
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAA	TAA	TAA
ND4	Length (bases/amino acid)	1380/459	1380/459	1380/459
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAG	TAG	TAG
ND5	Length (bases/amino acid)	1818/605	1818/605	1821/606
	Start codon	ATG(Met)	ATG(Met)	GTG(Val)
	Stop codon	TAA	TAA	TAA
Cytb	Length (bases/amino acid)	1143/380	1146/381	1143/380
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAA	TAA	TAA
NAD6 (L)	Length (bases/amino acid)	519/172	519/172	519/172
	Start codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop codon	TAG	TAG	TAG

Table 2
Length indicated in base pairs (bp) of penguin control regions, intergenic spacers and overlaps.

Region	*S. demersus* (current study)	*E. chrysocome* Watanabe et al. (2006)	*E. minor* Slack et al. (2003)
Control region	1758	1376	2040
tRNA-Phe/12S rRNA	–	–	–
12S Rrna/tRNA-Val	–	–	–
tRNA-Val/16S rRNA	–	–	–
16S Rrna/tRNA-Leu(UUR)	–	–	–
tRNA-Leu(UUR)/NAD1	5	4	5
NAD1/tRNA-Ile	2 overlap	2 overlap	2 overlap
tRNA-Ile/tRNA-Gln(L)	9	9	9
tRNA-Gln(L)/tRNA-Met	1 overlap	1 overlap	1 overlap
tRNA–Met/NAD2	–	–	8
NAD2/tRNA-Trp	2 overlap	2 overlap	2 overlap
tRNA-Trp/tRNA-Ala(L)	1	1	1
tRNA-Ala(L)/tRNA-Asn(L)	13	2	2
tRNA-Asn(L)/tRNA-Cys(L)	2	2	2
tRNA-Cys(L)/tRNA-Tyr(L)	1 overlap	1 overlap	1 overlap
tRNA-Tyr(L)/COX1	6	1	1
COX1/tRNA-Ser(UCN)(L)	9 overlap	9 overlap	9 overlap
tRNA-Ser(UCN)(L)/tRNA-Asp	6	5	4
tRNA-Asp/COX2	2	2	2
COX2/tRNA-Lys	1	1	1
tRNA-Lys/ATP8	1	1	1
ATP8/ATP6	10 overlap	10 overlap	10 overlap
ATP6/COX3	1 overlap	1 overlap	1 overlap
COX3/tRNA-Gly	–	–	–
tRNA-Gly/NAD3	–	–	–
NAD3/tRNA-Arg	4	4	4
tRNA-Arg/NAD4L	1	1	1
NAD4L/NAD4	7 overlap	7 overlap	7 overlap
NAD4/tRNA-His	1 overlap	1 overlap	1 overlap
tRNA-His/tRNA-Ser(AGY)	–	–	–
tRNA-Ser(AGY)/tRNA-Leu(CUN)	8	1 overlap	1 overlap
tRNA-Leu(CUN)/NAD5	–	–	–
NAD5/Cytb	6	7	7
Cytb/tRNA-Thr	4	4	3
tRNA-Thr/tRNA-Pro(L)	11	13	9
tRNA-Pro(L)/NAD6(L)	13	12	13
NAD6(L)/tRNA-Glu(L)	2	2	3
tRNA-Glu(L)/CR	–	–	–
CR/tRNA-Phe	–	–	–

8 bp spacer between tRNA–Met and NAD2 while the other two species have no spacer. The overlaps can be divided into at least four classes. The first class are those overlaps between H and L strand-encoded elements: 1 bp between tRNA-Gln(L)/tRNA–Met and 9 bp between COX1/tRNA-Ser(UCN)(L). Since different RNA transcripts are involved, these do not comprise genuine overlaps. The second class involve those overlaps on TAR stop codons: 2 bp overlap between a TAG stop codon in NAD2 and the start of tRNA-Trp; 1 bp overlap between ATP6 TAA stop codon and COX3. It may be that these are not true overlaps, but rather represent endonucleolytic cleavage sites producing incomplete stop codons (Ojala et al., 1981). The third class consists of overlaps between the coding sequences of PCGs: 10 bp between ATP8/ATP6 and 7 bp between NAD4L/NAD4 (in all birds). These overlaps are always associated with different reading frames, but more information is required regarding the generation and processing of mt protein-coding transcripts (Slack et al., 2003). Finally, the fourth class is made up of the remaining overlaps and involve unknown mechanisms: a 2 bp overlap between an AGG stop codon in NAD1 and the start of tRNA-Ile; a 1 bp overlap between tRNA-Cys(L) andtRNA-Tyr(L); a 1 bp overlap between tRNA-Ser(AGY) and tRNA-Leu(CUN) (not present in *S. demersus*).

3.4. Transfer RNA and rRNA genes

A total of 22 tRNAs are found interspersed in the mtDNA genome of *S. demersus* and range in size from 66 bp (tRNA-Ser(AGY) to 76 bp (tRNA-Trp and tRNA-Ser(UCN). The tRNAs include two tRNA-Leu and two tRNA-Ser. These tRNAs correspond to the standard set found in other metazoan mtDNAs. Most of the tRNAs could be folded into the canonical cloverleaf secondary structure with examples in Fig. 3A. As in vertebrates in general, the secondary structure of tRNA-Ser(AGY) lacks the DHU arm. Located between tRNA-Phe and tRNA-Val, the 12S rRNA gene of *S. demersus* was 980 bp (Fig. 3B), 4 and 5 bp longer than those described in *E. chrysocome* and *E. minor* respectively. The 16S rRNA gene (Fig. 3C), located between tRNA-Val and tRNA-Leu, was 1606 bp and was 2 bp shorter than *E. minor*, but 11 bp longer than *E. chrysocome*. Asakawa et al. (1995) suggested that a stem and loop structure around the 3′ end of 12 S rRNA and 16S rRNA could play an essential role in the protein synthesis and transcriptional regulation in mitochondria, respectively. A conserved 39 bp at the 3′ end of the 12S rRNA gene in *S. demersus* was also inferred to have a stable stem and loop structure (Fig. 3B) showing free energy of −17.7 kcal/mol. Although the 3′ end of 16 S rRNA of the penguins showed some sequence variation, they could still be folded into stem and loop structures (Fig. 3C). In eutherians, the L-strand origin of replication is usually located between tRNA-Asn and tRNA-Cys, but is missing in *S. demersus* with the two tRNAs separated by only 2 bp. The absence of an origin of replication at this position is consistent with other birds described (Mindell et al., 1998b). Desjardins and Morais (1991) proposed that it is possible for origin of L-strand replication to be initiated within the CR.

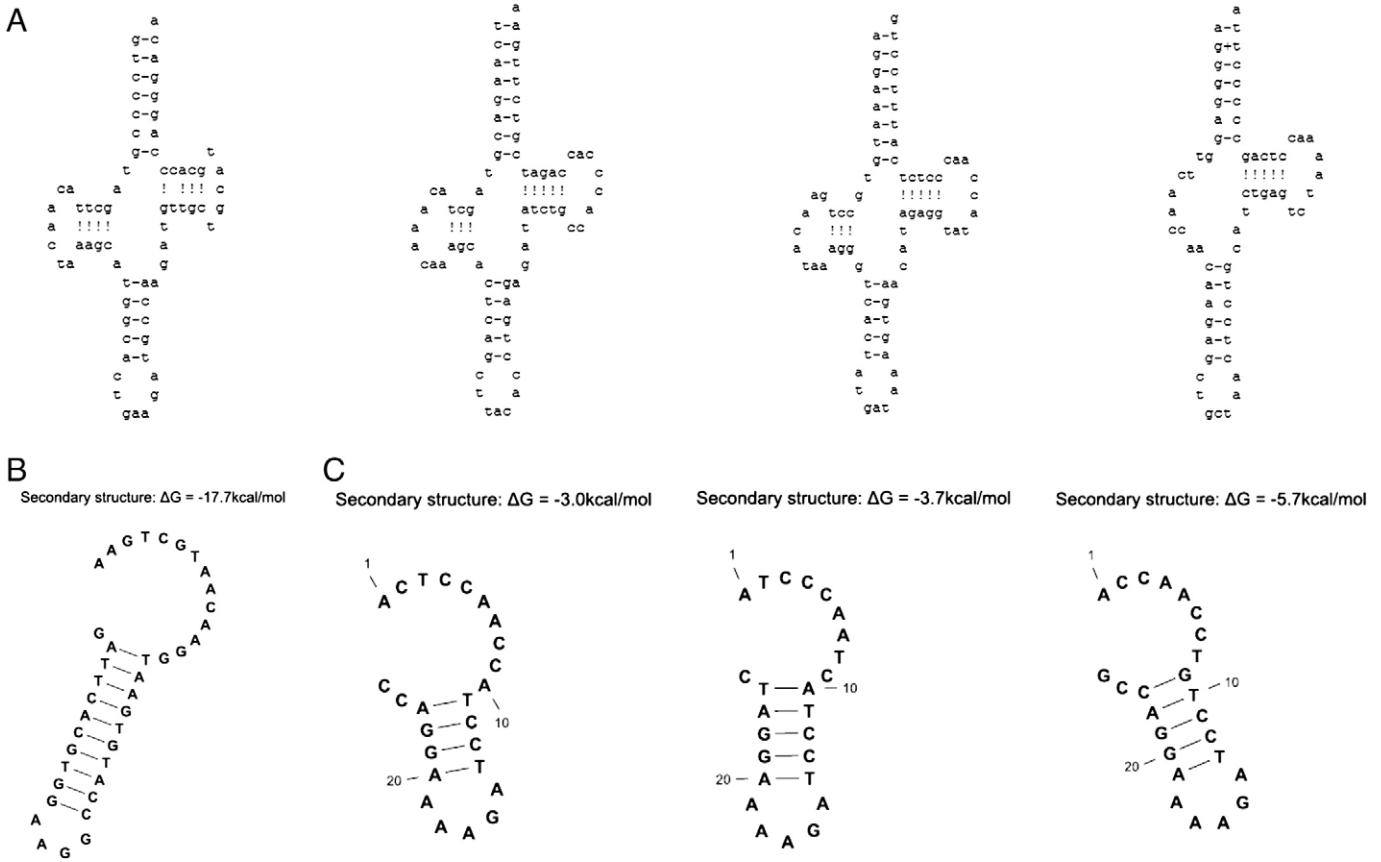

Fig. 3. Inferred secondary structures (A) Secondary structures of 4 tRNAs found in *S. demersus* namely; tRNA-Phe, tRNA-Val, tRNA-Ile, tRNA-Ser(AGY). (B) Secondary structure formed at the 3′ end of 12S rRNA found in *S. demersus*. (C) Secondary structures formed at the 3′ end of 16 S rRNA in *S. demersus*; *E. minor* and *E. chrysocome* respectively.

3.5. Control region

The mtDNA Control Region (CR) is responsible for transcription and replication of the mitochondrial genome (Taanman, 1999). As in the majority of birds for which data is available the CR for *S. demersus* is located between tRNA-Glu and tRNA-Phe. The CR of *S. demersus* is 1758 bp in length, which is longer than *E. chrysocome* (1376 bp), but shorter than what has been reported for *E. minor* (2040 bp). All three species contain two sets of similar repeats (TCGATACAYWTTACAY TTYWWYTTTCTCTAAAATTTCATTAABRYAYRATARCAACYCTTYGTTGCY ATCDYCTTTACTGTA and ACAACAAACAACAA) with varying copy numbers towards the 3′ end of the CR, accounting for the size variance. Conserved sequence blocks (CSB-1,-2, and-3) have been identified in the CR of several vertebrates and may be involved in the origin of H-strand replication (Walberg and Clayton, 1981). Only CSB-1 (TATTT GTTGAATGCTTGTTAGACATAA) could be identified in *S. demersus*. A cytosine string (CCCCCCCCTACCCCCC) located close to the 5′ end of the CR is similar to the motif observed in other avian species such as Struthioniformes, Galliformes and Falconiformes. The motif consists' of a G/C stem and a loop containing a TCCC motif that may be involved in H-strand termination (Ruokonen and Kvist, 2002). This motif in the CR has also been reported in African side-necked turtle (*Pelomedusa subrufa*) (Zardoya and Meyer, 1998). In chickens and lesser snow geese (*Ansercaerulescens caerulescens*) the motifs have the potential to form a stable hairpin structure (Quinn and Wilson, 1993). However, in the three penguins discussed here, the C-stretch is not followed by a G-stretch. Thus the repeat sequence is unable to form a hairpin secondary structure. Reasons behind conservation of the C-stretch is still unknown and the role of this sequence is currently unknown (Ruokonen and Kvist, 2002). The termination-associated sequence motif TATAT was identified 33 bp downstream from the C-stretch in *S. demersus*, but was not present in the other two penguins. The termination-associated sequence motif TACAT, immediately preceding the TATA motif in *S. demersus* is present in all three species. The highly conserved bird similarity box (CACTGATGCACTTTG) was identified approximately 821 bp downstream from the C-stretch in all three penguins. The high level of sequence conservation suggests that the bird similarity box may play a key role in the replication and transcription of the mitochondrial genome in Aves (Bing et al., 2006).

In summary, this is the first report of the complete nucleotide sequence for the mitochondrial genome of the African penguin, *S. demersus*. The *S. demersus* mtDNA genome is very similar, both in composition and length to both the *E. chrysocome* and *E. minor* genomes. These results can be subsequently used to provide information for penguin phylogenetic studies and insights into the evolution of genomes.

Conflict of interest

The authors have no conflict of interest.

Acknowledgments

We thank the South African Association for Marine Biological Research, KwaZulu-Natal South Africa for the penquin sample. Work at the National Zoological Gardens of South Africa (NZG) was partially supported by the National Research Foundation (NRF) grant (grant number 79732), the Society, Ecosystems and Change (Seachange).

References

Asakawa, S., Himeno, H., Miuraand, K., Watanabe, K., 1995. Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics 140, 1047–1060.

Avise, J.C., 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall, New York.

Baker, A.J., Pereira, S.J., Haddrath, O.P., Edge, K., 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proc. R. Soc. B 273, 11–17.

Bing, X., Fei, M.A., Yi, S., Qing-Wei, L., 2006. Comparative analysis of complete mitochondrial DNA control region of four species of strigiformes. Acta Genet. Sin. 33 (11), 965–974.

Brown, W.M., 1983. Evolution of animal mitochondrial DNA. In: Nei, M., Koehn, R.K. (Eds.), Evolution of Genes and Proteins. Sinauer, Sunderland, Massachusetts, pp. 62–88.

Cao, S., Wu, B., Yan, P., Hu, Y., Su, X., Jiang, Z., 2006. Complete nucleotide sequences and gene organization of mitochondrial genome of *Bufo gargarizans*. Mitochondrion 6, 186–193.

Desjardins, P., Morais, R., 1990. Sequence and gene organization of the chicken mitochondrial genome. J. Mol. Biol. 212, 599–634.

Desjardins, P., Morais, R., 1991. Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J. Mol. Evol. 32, 153–161.

Gibb, G.C., Kardailsky, O., Kimball, R.T., Braun, E.L., Penny, D., 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol. 24 (1), 269–280.

Kan, X.Z., et al., 2010. Phylogeny of major lineages of galliform birds (Aves: Galliformes) based on complete mitochondrial genomes. Genet. Mol. Res. 9 (2), 1204–1216.

Labuschagne, C., van Wyk, A.M., Kotze, A., Grobler, P., Dalton, D.L., 2013. Isolation and characterization of species-specific microsatellites loci in African penguin (*Spheniscus demersus*). Conserv. Genet. Resour. 5, 169–171.

Laslett, D., Canbäck, B., 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172–175.

Lei, R., et al., 2010. Complete sequence and genome organization of the mitochondrial genome for Hubbard's sportive lemur (*Lepilemur hubbardorum*). Gene 464, 44–49.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998a. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Mol. Biol. Evol. 15, 1568–1571.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998b. Mutiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. U. S. A. 95, 10693–10697.

Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470–474.

Quinn, T.W., Wilson, A.C., 1993. Sequence evolution in and around the mitochondrial control region in birds. J. Mol. Evol. 37, 417–425.

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

Russell, R.D., Beckenbach, A.T., 2008. Recording of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. J. Mol. Evol. 67, 682–695.

Sammler, S., Bleidorn, C., Tiedemann, R., 2011. Full mitochondrial genome sequence of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. BMC Genomics 12, 35.

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 (1–2), 43–52.

Taanman, J.-W., 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta 1410, 103–123.

Walberg, M.W., Clayton, D.A., 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. 9, 5411–5421.

Watanabe, M., et al., 2006. New candidate species most closely related to penguins. Gene 378, 65–73.

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

Wu, X.B., Wang, Y.Q., Zhou, K.Y., Zhu, W.Q., Nie, J.S., Wang, Z.L., 2003. Complete mitochondrial DNA sequence of Chinese alligator *Alligator sinensis* and phylogeny of crocodiles. Chin. Sci. Bull. 48 (19), 2050–2054.

Zardoya, R., Meyer, A., 1998. Cloning and characterization of a microsatellite in the mitochondrial control region of the African side-necked turtle, *Pelomedusa subrufa*. Gene 216, 149–153.

Zhou, Y., Zhang, Jia-Y., Zheng, Rong-Q., Yu, Bao-G., Yang, G., 2009. Complete nucleotide sequence and gene organization of the mitochondrial genome of *Paa spinosa* (Anura: Ranoidae). Gene 447, 86–96.