Comparative genomic analysis identifies small open reading frames (sORFs) with peptide-encoding features in avian 16S rDNA

Mathieu Mortz, Cyril Degletagne, Caroline Romestaing, Claude Duchamp

Comparative genomic analysis identifies Small Open Reading Frames (sORFs) with peptide-encoding features in avian 16S rDNA

Mathieu Mortz, Cyril Dégletagne, Caroline Romestaing, Claude Duchamp*

Université de Lyon, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, UMR 5023 CNRS, Université Claude Bernard Lyon 1, ENTPE, Villeurbanne cedex, France

*Corresponding author: *Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, UMR 5023 CNRS, Université Claude Bernard Lyon 1, ENTPE, 43 Bvd 11 Novembre 1918, F-69622 Villeurbanne cedex, France.*

Email: claude.duchamp@univ-lyon.fr

Short title: Putatively functional sORFs in avian 16S rDNA

Abstract

The mitochondrial genome (mt-DNA) functional repertoire has recently been enriched in mammals by the identification of functional small open reading frames (sORFs) embedded in ribosomal DNAs. Through comparative genomic analyses the presence of putatively functional sORFs was investigated in birds. Alignment of available avian mt-DNA sequences revealed highly conserved regions containing four putative sORFs that presented low insertion/deletion polymorphism rate (<0.1%) and preserved in frame start/stop codons in >80% of species. Detected sORFs included avian homologs of human Humanin and Short-Humanin-Like-Peptide 6 and two new sORFs not yet described in mammals. The amino-acid sequences of the four putative encoded peptides were strongly conserved among birds, with amino-acid p-distances (5.6 to 25.4%) similar to those calculated for typical avian mt-DNA-encoded proteins (14.8%). Conservation resulted from either drastic conservation of the nucleotide sequence or negative selection pressure. These data extend to birds the possibility that mitochondrial rDNA may encode small bioactive peptides.

Keywords: Mitochondrial genome; Ribosomal RNAs; Humanin; Short-Humanin-Like Peptides; SHLP6

1. Introduction

Mitochondria are cell organelles mainly involved in the production of ATP required to sustain biological exergonic processes (1). ATP synthesis mostly occurs within the oxidative phosphorylation (OXPHOS) system that is composed of five protein complexes combining subunits encoded by nuclear (nuc-DNA) and mitochondrial (mt-DNA) genomes (2-3). In animals, mt-DNA usually contains 37 genes that encode 22 tRNAs, 2 rRNAs (12S & 16S) and 13 mitochondrial protein-coding genes (mt-PCGs) that complement the subunits encoded by nuc-DNA to constitute the OXPHOS system (4).

This traditional mt-DNA description has been questioned in the last decade with the description in various distantly related multicellular organisms of highly conserved small open reading frames (sORFs), scattered within mt-DNA and showing characteristics suggesting encoding function (5). In human mt-DNA, eight sORFs have already been identified as effectively encoding small peptides and called Humanin (Hn) (6-7), MOTS-c (8) or SHLP1-6 (9). These peptides encoded by sORFs of different sizes (45-90 pb) embedded on both strands of the 12S (MOTS-c) or 16S (Hn, SHLP1-6) rDNAs are obtained after mitochondrial or cytosolic translation (10). Translated products are detected in the bloodstream and exert a large spectrum of biological actions including anti-apoptotic and metabolic effects (11). Owing to the large number of identified peptides and their biological roles, a novel concept of mitochondria-derived peptide-signalling is slowly emerging in recent years (10, 12). To date studies focused on human and rodent laboratory models, suggesting that this concept may be extended at least to the class of mammals. However the existence of a dual encoding role (rRNA and small peptides) for mitochondrial rDNAs has rarely been explored in other classes of animals. Only one study, using data mining approaches, reported that the humanin gene was conserved in 10 bird species while it was pseudogenized in many vertebrates (13). The common presence of the humanin gene in primates, some rodents and birds (13) warrants investigating more deeply the putative presence of mt-DNA-encoded peptides in birds.

Here, we propose an original *in silico* blind approach to test the peptide-encoding potential of 12S and 16S rDNAs in the highly diversified class of birds through the detection of sORFs. Birds together with mammals exhibit endothermy that requires a tachymetabolism that might necessitate preserving metabolically active signals. Functional sORFs should a priori be found in highly conserved regions of 12S/16S rDNAs and should also fulfil characteristics of functional coding regions such as reading frame maintenance and amino-acid sequence conservation. Our bioinformatics analysis therefore included stringent filters (14) to avoid the detection of false-positive sORFs.

2. Material and Methods

2.1. Collection of mt-DNA sequences and alignment

All available (at the time of the study) complete avian mt-DNA sequences (RefSeq) were extracted from GenBank (552 species). Sampled species represented all avian orders in various proportions (42% Passeriformes, 11% Galliformes, 5% Anseriformes, 5% Pelecaniformes, 5% Gruiformes, 5% Psittaciformes, 5% Charadriiformes, 3% Collumbiformes, 3% Apodiformes, and 0.2-2% each for all other orders). Collected sequences of avian mt-DNA were aligned using MUSCLE (15) through MEGA07 software (16).

2.2. Identification of highly conserved rDNA regions

Highly conserved rDNA regions were identified by comparing region conservation to the average conservation of the rDNA alignment. Conservation (C) corresponds to the proportion of invariable sites (C=1-(S/L)), where S is the number of variable sites and L is the net number of analysed positions. Conservation threshold was fixed to keep the rDNA regions that are conserved at least 10% more than the rest of the rDNA gene (rDNA region

(C) > entire rDNA (C) + 0.1). The window width of rDNA region was determined as the minimum width that allowed a region to be significantly conserved at α=0.05 (under the hypergeometric distribution). To select only highly conserved regions, we kept detected regions with $\alpha<10^{-6}$. Calculations (17) were made using DNAsp v5.10.01 (18).

2.3. sORFs detection

sORFs were detected by submitting the consensus sequences (threshold 25%, considering gaps) of highly conserved regions to the NCBI bioinformatics tool ORFFinder. Both heavy and light mt-DNA strands were screened and potential translation was addressed using mitochondrial or standard genetic codes. To avoid the detection of an excessive number of potential interesting sequences, the minimal length of detected sORFs was set at 45 nucleotides (https://www.ncbi.nlm.nih.gov/orffinder/).

2.4. Peptide-encoding features and sORF selection process.

A multiallelic estimation of the level of insertion/deletion (InDel) polymorphism (considering all insertion/deletion events) was made for the concatenate of the thirteen mt-PCGs (in the order: ATP6/8, COX1/2/3, CytB, ND1/2/3/4/4L/5/6), for the two 12S and 16S rDNAs and for all highly conserved sORFs detected. This step was defined to select sORFs with a level of InDel polymorphism close to or lower than that measured for all mt-PCGs with the assumption that coding sequences should have low InDel polymorphism to maintain reading frame and conserve the translated a-a sequence. We then verified, for each of the 552 species studied, the effective presence of in frame (compared with initially detected consensus sORF) start codons (“ATG” for standard genetic code; “ATG” or “ATA” for mitochondrial genetic code) and stop codons (“TAA”, “TAG” or “TGA” for standard genetic code; “TAA”, “TAG”, “AGA” or “AGG” for mitochondrial genetic code). This enabled us to estimate a percentage of sORF preservation that corresponds to the proportion of species with a given putative encoding sORFs ≥45 nucleotides flanked by in frame start/stop codons. A high degree of preservation is expected for functional and biologically relevant peptide

encoding sORFs. A stringent threshold of ≥80% of preservation among species led to select only a limited number of sORFs that were compared with human homologous mt-DNA regions. Comparison was performed by aligning avian 16S consensus rDNA with human 16S rDNA reference sequence (Homo sapiens mitochondrion complete genome, NCBI Reference Sequence NC_012920.1, Position: 1671:3229). We also estimated for each of the remaining selected sORFs the non-synonymous substitution rate (number of non-synonymous substitutions / non-synonymous sites, Ka), the synonymous substitution rate (number of synonymous substitutions / synonymous sites, Ks) (19-25) and both nucleotide and amino-acid p-distances. Values were compared with those calculated for avian mt-PCGs. All calculations were made through MEGA07 (16) and DNAsp v5.10.01 (18).

2.5. RT-PCR analysis of sORF expression.

sORF functionality was assessed by the detection of targeted transcripts using reverse transcription-polymerase chain reaction (RT-PCR). Total RNA was extracted with Trizol from either pectoralis muscle sample obtained from a 6-wk old Muscovy duckling (*Cairina moschata*). After digestion of RNA extracts with DNase I (Thermo Fisher Scientific), mRNA were purified using a Dynabeads® mRNA purification kit (Thermo Fisher Scientific) and quantified using a Qubit™ Fluorometer (Thermo Fisher Scientific). Reverse transcription (RT) of 1 µg RNA was performed using polyT as primer and 200 U MMLV-RT (InVitrogen) according to the manufacturer protocol. PCR was then performed in a total volume of 20 µL using platinium Taq polymerase (Invitrogen) and a Hybaid thermocycler (Thermo Fisher Scientific). After an initial denaturation at 94°C for 3 minutes, 30 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 30 s and extension at 72°C for 60 s were used. A final extension was performed at 72°C for 10 min. Forward and reverse primers were designed from muscovy duck 16S rDNA sequence (GenBank Accession: NC_010965.1) and were: av-Hn 5'- CCT AAC CGT GCA AAG GTA GC-3' and 5'- AGT TCC ACA GGG TCT TCT CG -3' (product size 155 bp); av-scO2 5'- ATA GAA ACC GAC CTG GAT TG -3' and 5'- GTT TAC GAC CTC GAT GTT GG -3' (product size 127 bp). Actin was used as a positive control for

mRNA extracted from muscle 5'-GAC GAG GCC CAG AGC AAG AGA-3' and 5'-GGG TGT TGA AGG TCT CAA ACA-3' (product size 225 bp). A positive control of mt-DNA transcription was also used: av-Cox1 5'-AAC TAG GCC AAC CAG GGA CT-3' and 5'-AGA TGA GGC GAG CAG TAG GA-3' (product size 230 bp). The amplified products were separated according to their size on 2% agarose gel stained with ethidium bromide and images were acquired with a Kodak Digital Science™ DC120 Camera.

3. Results and Discussion

3.1. Identification of highly conserved regions & detection of avian sORFs

The conservation pattern of avian rDNAs was assessed by calculating nucleotide divergences for all positions without insertion/deletion event along the 12S and 16S rDNAs. As shown in Fig.1, the average nucleotide divergence among birds of a 45 nucleotides sliding window was heterogeneously distributed along 12S and 16S rDNAs. In both 12S (Fig.1A) and 16S rDNAs (Fig.1B), divergence was heterogeneously distributed with regions of relatively high divergence (>15%) and others with very low divergence (<5%), suggesting that some regions are more functionally constrained among bird species. Heterogeneous conservation along 12S/16S rDNA with strongly conserved positions has already been described in several prokaryotic and eukaryotic taxa (27, 28), but present data are the first to focus specifically on birds.

Heterogeneous conservation of rDNA sequence is not surprising as evolution of rDNA depends on the secondary structure of encoded rRNA, with fast- and slow-evolving positions in loop-encoding and stem-encoding regions respectively (29). Specific peptide-encoding sORF in slow-evolving regions could constitute an additional constraint considerably affecting the conservation of some specific portions of rDNA. Accordingly, our analysis on bird sequences revealed three hotspots of remarkably small divergences in either 12S (2.1 ±

186 0.9% at Pos. 508 (window 485-529)) or 16S rDNA (1.4 ± 1.0% at Pos. 840 (window 819-863)
187 or 0.08 ± 0.03% Pos. 1303, (window 1282-1326)) (Fig.1). Assuming that important peptide
188 encoding rDNA regions should be strongly conserved during bird evolution, we looked for
189 highly significantly ($p<10^{-6}$) conserved regions in 12S/16S rDNAs. Calculations (see
190 Methods) led to identify 5 significantly ($p<0.05$) conserved regions and 3 of them, located
191 either in 12S or 16S rDNAs, were highly significantly ($p<10^{-6}$) conserved among bird species
192 and logically contained the three hotspots of drastic conservation (Fig.1).

193 Using the avian consensus sequences of the identified highly conserved regions,
194 sORFs were then detected with ORFFinder. We searched for potential coding sequences
195 ≥15 a-a (≥45 nucleotides) with reference to vertebrate mitochondrial genetic code and/or
196 standard genetic code, according to the features noticed in mammals for the peptide-
197 encoding sORFs already identified (5, 6-9, 14). This strategy enabled us to detect 17 sORFs
198 embedded in either 12S (5 sORFs, Fig.1A) (Table S1) or 16S (12 sORFs, Fig.1B) (Table
199 S2). sORFs were numbered according to their position in the sequences, 1 to 5 in 12S and 6
200 to 17 in 16S and labelled as being detected by referring to either mitochondrial genetic code
201 (e.g. 1m/2m...) or standard genetic code (e.g. 1s/7s...). Two sORFs were detected by
202 referring to both mitochondrial and standard genetic codes (1m/1s/10m/10s) and were each
203 considered as distinct sORFs (mitochondrial vs cytosolic forms) for the following analyses,
204 because of genetic code-dependent variations in i) sORF length related to differences in
205 sORF initiation and stop codons positions and ii) in the composition of translated a-a
206 sequences.

207 *3.2. Low level of InDel polymorphism in avian sORFs?*

208 The mere fact of detecting sORFs in conserved regions does not mean that they are
209 functional and specifically conserved during evolution (30, 31), especially since avian rDNA
210 sequences contain a high percentage of GC both in 12S (49.5%) and 16S (45.1%) that
211 would likely favour the random presence of sORFs in non-coding sequences. However, a
212 direct consequence of the existence of functional peptide encoding regions is a weak level of

insertion / deletion (InDel) events to maintain reading frame and thus to conserve the translated a-a sequence. This feature can be verified for avian mt-DNA genes by measuring the InDel polymorphism rate (InDel Pi) for rDNAs and for the concatenate of all mt-PCGs. The InDel Pi (/100 sites) was more than 20 times lower for mt-PCGs (0.16%) than that measured for 12S (3.68%) or 16S (4.6%) rDNAs (Fig.2).

Within the detected conserved sORFs, nine sORFs presented InDel Pi in the same order of magnitude (ranging from 0.96 to 3.55%) as those measured for the entire 12S or 16S rDNAs suggesting the absence of specific pressure during evolution to maintain the reading frame by comparison with the rDNA gene in which they are included. By contrast, seven sORFs (2_m, 9_m, 10_{ms}, 11_s, 15_m, 16_s, 17_m) presented very low InDel Pi (ranging from 0.007 to 0.094%) that were even lower than that measured for the concatenate of avian mt-PCGs (Fig.2). These extremely low InDel Pi are consistent with the maintenance of putative reading frames, as would be expected in peptide/protein-coding regions. Present data therefore suggest that the seven sORFs detected in the consensus avian 16S rDNA might encode peptides that impose functional constraints on sequences during evolution by comparison with the rest of the gene in which they are included.

3.3. In frame sORF preservation among bird species?

To provide additional arguments in favour of their putative peptide-encoding role, we focused on these seven detected sORFs and verified their effective presence in all 552 bird species. For each sORF in every species, we looked for the presence of both start and stop codons in frame with the initially detected consensus sORF and flanking a region of ≥45 nucleotides. We then calculated the percentage of sORF preservation (Fig.3) that corresponds to the proportion of bird species meeting those criteria and kept sORFs preserved in more than 80% of studied bird species. This high level of preservation was chosen to keep sORFs with a significant biological role potentially imposing strong conservation with evolution.

As shown in Fig.3, estimations of sORF preservation varied among the seven selected sORFs indicating that neither the low nucleotide diversity (Fig.1) nor the low InDel polymorphism rate (Fig.2) of the sORFs in the consensus sequence would ensure their preservation in all bird species. For the last 12S sORF detected in the consensus sequence and that fulfilled previous filters (2_m), only 41.3% of bird species preserved a ≥45 nucleotides sORF. Similarly, despite strong nucleotide sequence conservation, sORF 9_m and 15_m were preserved in 66.3% and 29.9% of birds, respectively. Only a limited number of sORFs were preserved in more than 80% of studied species and they were all found in 16S rDNA: 10_m (82.1%), 10_s (81.3%), 11_s (91.3%), 16_s (97.8%) and 17_m (99.1%). Such high degree of preservation is indicative of functional constraints to maintain the sequence of specific sORFs in most bird species. The presence of sORFs with robust peptide-encoding features exclusively in avian 16S rDNA is consistent with the observations in mammals in which identified encoding sORFs were mostly found within 16S rDNA (n=7) (6, 7, 9) while only one encoding sORF was detected in 12S (8).

3.4. Are avian sORFs potentially coding for orthologs of peptides already identified in humans?

The consensus sequences of the four remaining avian 16S sORFs were aligned with human 16S rDNA sequence (NC_012920.1, Pos.1671:3229). Both avian consensus and human sORF homologous regions were translated into a-a sequences using appropriate genetic codes and compared (Table 1). Among the four remaining sORFs, we identified putative avian orthologs of humanin (sORF $10_{ms:}$ av-Hn) and SHLP6 ((sORF 16_s: av-SHLP6) that were encoded by the strand (+) of avian 16S rDNA. The deduced sequence of av-Hn consensus peptide (obtained after either mitochondrial or cytosolic translation) showed 58% identity (14 of the 24 a-a were identical) with the cytosolic form of human Hn. Present data confirm and extend the previous finding of a potentially functional humanin gene in ten bird species (13) by documenting the conservation of av-Hn in >80% of the 552 bird species in which mt-DNA sequence was available. There was an even greater degree of a-a

conservation through evolution for SHLP6 as the human and avian proteins showed 90% identity (18/20 identical a-a) (Table 1).

The presence of conserved orthologs for humanin and SHLP6 in birds and mammals suggests that these peptides exhibit essential biological roles. In mammals, humanin was shown to reduce apoptosis (32, 33) and to favour cardioprotection (34), while the various SHLPs were shown to act as age-dependent regulators of apoptosis, insulin sensitivity and inflammatory markers (9). The essential actions of these peptides as potent cell-survival and metabolic factors might contribute, at least in part, to the necessity to preserve their a-a composition and functional a-a motifs (Table 1) during evolution. It remains to be shown that these biological roles have been preserved in birds.

The other two sORFs (11_s, 17_m) identified in our analysis were both putatively encoded by the strand (-) of avian 16S rDNA and did not match with any characterized human encoding sORFs. These new identified sORFs were called av-scO1 (11_s) and av-scO2 (17_m) for “avian strongly conserved sORFs 1 and 2”. The presence of putative encoding sORFs in the non-coding strand (-) of avian 16S rDNA is in agreement with the presence of sORFs encoding SHLPs (SHLP1 to 5) in the non-coding strand (-) of human 16S rDNA (9). Interestingly, despite a global overlap with av-Hn (pos. 886-960, strand (+)) and thus a same substitution rate in their common nucleotide area (pos. 913-960), no homologous sORF was identified in human for av-scO1. The translated human a-a sequence would result in a a-a sequence with only 24% identity (4/17 a-a identical) with the avian putative peptide. For av-scO2, the potential human peptide despite being shorter would show 94% identity with the avian form (16/17 a-a identical). These observations suggest that av-scO1 appears to be specific to bird species while av-scO2 would be a more ubiquitous sORF highly conserved at least in endothermic vertebrates. The experimental confirmation that these sORFs are functional and encode for biologically active peptides now remains to be appropriately verified. To provide additional arguments concerning the potential encoding function of detected sORFs we analysed the nucleotide composition around the start codon to detect

potential Kozak sequence (GCC(G/A)CCAUGG) that represent translation initiation signals (TIS) facilitating the recognition of AUG codon by ribosomal subunits (35). No strong robust Kozak-like sequences could be found in the detected sORFs. When focusing on the two main sites (3(G/A) and +4G) associated with high translation efficiency in mammals (36), we only noted the presence of +4G for av-Hn and av-scO2. Nevertheless, the absence of characteristic TIS does not mean that selected sORFs could not encode peptides on account of the fact that only 40% of mammalian mRNAs present highly efficient TIS and 10% of them do not present the characteristic nucleotides -3(G/A) and +4G (37).

Since our large scale blind analysis successfully identified putative avian orthologs of peptides already identified in mammals, it gave some confidence to tackle the characterization of the peptides encoded by the new sORFs. To strengthen our analyses, we harnessed available bird transcriptomic data (NCBI GEO DataSets) to determine whether the putative sORF are transcribed. Most interestingly, we found sequences corresponding to the four selected sORFs in mRNA libraries obtained from several tissues of *Gallus gallus* (38) or *Anser cygnoides* (39) including heart, breast muscle or adipose tissue. Note that mRNA libraries were prepared for sequencing using TruSeq stranded mRNA sample preparation guide or the Ribo-Zero^{TM} kit to remove rRNA. Although the possibility that transcripts containing the sORFs encoded by the strand(+) could be related to traces of 16S rRNA not removed by mRNA purification cannot be fully excluded, the finding of transcripts encoded by the stand(-) such as av-scO1 and 2 is not expected unless they relate to specific transcriptions of these regions (9). To further strengthen our in silico findings we performed preliminary investigations of potential transcription using RT-PCR detection of one sORF of each strand (av-Hn and av-scO1). As shown in Fig.4, RT-PCR amplicons of expected size could be found using purified mRNAs extracted from duck pectoralis muscle. When RT was omitted, no amplicon was detected showing the specific amplification of poly-adenylated transcripts. Although these preliminary observations are promising signatures of sORF functionality, the presence of translated peptides now requires to be investigated.

3.5. Negative selection pressure for avian sORFs?

The amino acid composition conservation of a peptide / protein coding gene among species is known to result from a strong global nucleotide conservation, but also from an increase of the inconsequential rate of synonymous (syn.) substitutions (number of syn. substitution / number of syn. sites, Ks) at the expense of non-syn. substitutions rate (number of non-syn. substitution / number of non-syn. sites, Ka). Thus, negative selection pressure (Ka/Ks ratio <1) highlighted when comparing nucleotide sequence of different species is generally the consequence of the presence of common peptide / protein coding region, which requires to conserve the amino acid composition during evolution (40). To assess the type of selection pressure underwent by the four selected 16S sORFs within bird species, both Ka and Ks were estimated for each pairwise comparison of species and compared with that underwent by the 13 mt-PCGs. Ka values were plotted as a function of Ks values for each sORF and for the concatenate of avian mt-PCGs. With 552 bird species to compare, this represented 152 352 pairwise comparisons per plot. To simplify the representation, only the regression lines of each plot were drawn (Fig.5) and their respective slope (Ka/Ks) and regression coefficient (R^2) were given. Results indicated that except for av-scO2 that appeared to be under neutral selection (Ka/Ks slope = 0.94), av-scO1 (Ka/Ks slope = 0.32) and the putative orthologs of humanin (Ka/Ks slope = 0.25 (mt.) / 0.22 (std.)) and SHLP6 (Ka/Ks slope = 0.36) were all under a strong negative selection pressure during bird evolution. Such negative selection pressure is indicative of the need to conserve the amino acid composition of the putative encoded peptides.

As shown in Fig.5, the concatenate of proteins encoded by avian mt-PCGs appeared under an even higher (Ka/Ks slope = 0.10) negative selection than the selected avian sORFs suggesting that the conservation of mt-proteins among bird species would be accordingly higher. To verify this point, we calculated the amino acid conservation (p-distances) among birds for the peptides/proteins encoded by the four sORFs and mt-PCGs. As shown in Fig. 5A, similar or greater conservations of a-a composition were unexpectedly found for the

translated peptide sequences of av-Hn (15.4 ± 4.8% (mt.); 16.9 ± 4.6% (std.)), av-SHLP6 (6.6 ± 2.0%), av-scO1 (25.4 ± 4.9%) and av-scO2 (5.6 ± 2.0%) as compared with the concatenate of all mt-PCGs (14.8 ± 0.3%) (Fig.6A). Despite overall conservation of sequences, there were small substitutions in amino-acid composition that could affect the functionality of the encoded peptide depending on the region concerned. It is difficult at this stage to conclude about their impact on the functionality of the encoded peptides without molecular validation of peptide function at the amino-acid level. However, the conservation of the two putatively functional motifs found in av-Hn ("SEIDL") or av-scO2 ("TDLD") was analysed and compared with the global p-distance measured for the entire sORF. We found a remarkable conservation of the functional motif of av-scO2 TDLD (1.0 ± 0.2%) that underwent a stronger constraint than the entire av-scO2 sORF (5.6 ± 2.0%), suggesting a putative important role in the peptide functionality. By contrast, the av-Hn functional motif SEIDL was not more conserved (21.7 ± 0.17% (both mt. and std.)) than the entire sORF (15.4 ± 4.8% (mt.); 16.9 ± 4.6% (std.)). This was related to a strong positive selection underwent by the first a-a of the SEIDL motif (mean Ks: 3.7%; mean Ka: 25.6%), while the last four positions were less variable (10.7 ± 4.7%). Thorough analysis revealed the predominance of only two amino-acids at the first position (the fourteenth position when considering entire av-Hn peptide sequence) of the motif, with 291 species (53%) presenting a serine (Ser14) and 228 species (41%) presenting a glycine (Gly14) at this position. Interestingly, a Gly14-Hn variant was also characterized in mammals, and was shown to have a 1000-fold higher neuroprotective activity than the Ser14-Hn variant (41). The non-synonymous mutations observed in birds at the 14th position of av-Hn might thus reflect an adaptation of av-Hn functionality in birds that encode a Gly14-Hn but additional experimental verifications of these aspects are needed. Another functional site described in the literature for Hn was the cysteine at the eighth position of the peptide (Cys8), observed in both avian consensus and human peptide sequences. The presence of a cysteine at this position was essential to the anti-apoptotic function of the peptide (42). This amino acid was found to be present in nearly all 552 bird species studied, except one (*Turdus philomelos*) that presented

a frameshift that modified the encoding amino-acid at this position. More investigations are needed to verify if this InDel event indeed occurred or if it is only the consequence of a sequencing error.

Interestingly, the four putatively encoded peptides showed greater amino-acid conservation than those measured for known mt-PCGs in contrast with calculated Ka/Ks values (Fig.5). However, negative pressure intensity not only depends on the functional importance of the sequence of the encoded protein in studied species but also on the global substitution rate observed between nucleotide sequences (43-45). We therefore plotted Ka/Ks ratio as a function of nucleotide conservation for the known or putative coding sequences in birds (Fig.6B). There was a global negative relation between negative pressure intensity and the nucleotide substitution rate estimated by measuring the nucleotide p-distance for each studied regions (Fig.6B). For similar nucleotide conservation, for instance with av-SHLP6 and av-scO2, Ka/Ks values could be very different. This might be due to the fact that an increase of the global substitution rate in protein coding regions will mainly affect synonymous sites to conserve a-a composition. Consequently, fast-evolving coding genes show higher Ks values and consequently lower Ka/Ks ratio than slow-evolving genes, regardless of the functional importance of the encoded a-a (43). Therefore, av-scO2, drastically conserved among bird species (nuc. p-distance= 2.0±0.9%) and resulting in a strong peptide conservation (5.6±2.0%), could be qualified as a slow-evolving coding gene as compared with the fast-evolving mt-PCGs (nuc. p-distance= 20.8±0.2%) showing higher protein divergence (14.8±0.3%). Thus, the apparent neutral selection observed for av-scO2 (Ka/Ks= 0.94) in Fig.5 was the consequence of extremely low values of both Ka (mean Ka= 3.2%) and Ks (mean Ks= 2.3%), as compared with the values calculated for the concatenate of all mt-PCGs (mean Ka= 8.3%; mean Ks= 55%). It was not related to the importance of av-scO2 a-a composition. Altogether, present results indicate that the strong conservation of the deduced a-a sequence for the four selected 16S sORFs, that was similar to that calculated

401 for avian mt-PCGs, was the result of drastic nucleotide conservation (av-SHLP6 and av-
402 scO2) or strong negative selection pressure (av-Hn and av-scO1) during bird evolution.

403 3.6. Are there only four putatively encoding sORFs in avian 12S and 16S rDNA?

404 Our cautious approach to detect sORFs verifying peptide-encoding features after
405 applying stringent filters aimed at minimizing the risk of selecting false-positive sequences.
406 The down side of this drastic selection is that some of the rejected sORFs, not fulfilling all
407 criteria, could anyway encode peptides. Filtering on InDel polymorphism was relevant
408 because it identified a distinct sORF population presenting very low InDel rates, in
409 agreement with the need to maintain reading frames specific to peptide/protein encoding
410 regions. Filtering on the presence of both start and stop codons in frame with initially
411 detected consensus sequences should however be moderated. Indeed, the conservative
412 vision which attributes to “AUG” (and “AUA” in mitochondria) the exclusive initiation of
413 translation is debated (46). As an example, when regarding all the avian mt-PCGs collected
414 for this study, we noticed a non-traditional initiation codon (other than “AUG” or “AUA”) in
415 15.5% of bird sequences. In a similar way, the condition of in frame start codon is not verified
416 for SHLP1 and SHLP3 in the 16S rDNA of Sprague Dawley rats (Table S3), while the
417 peptides are detected in blood and SHLP3 exhibits biological effects (9). We therefore
418 analyzed the selection pressure underwent by the three sORFs discarded due to their low
419 percentage of “sORF preservation” (2m, 9m, 15m). Interestingly, we found similar Ka/Ks
420 values to those found for the four selected sORFs that fulfilled all criteria, with a strong
421 negative selection for 12S-sORF 2m (Ka/Ks slope: 0.30) and 16S-sORF 9m (Ka/Ks slope:
422 0.15) allowing a good conservation of amino-acid composition (7.8 ± 4.7% and 8.5 ± 3.8%,
423 respectively) (Table S1/S2). The 16S-sORF 15m underwent a less potent negative selection
424 (Ka/Ks slope: 0.65), but the strong conservation of its nucleotide composition (7.9 ± 2.3%)
425 enabled a good conservation of amino-acid composition (11.8 ± 4.6%) (Table S2), as was
426 the case for av-SHLP6 and av-scO2. These 3 sORFs, even if they are not as robust
427 candidates as the four sORFs selected by our stringent cut-off selection, might also encode

peptide in bird species. It follows that additional molecular and biological approaches are requested to overcome the limits of bioinformatics analysis (47) and draw up an exhaustive list of all the peptides effectively encoded by mitochondrial rDNAs.

4. Conclusion

A large scale comparative genomic analysis of available avian mt-DNA sequences (552 species) showed that in highly conserved regions of avian 16S rDNA, four sORFs exhibited peptide-encoding features comparable to those measured for all avian mt-PCGs, with a low InDel polymorphism rate (<0.1%), a high preservation of both start and stop codons (present in more than 80% of species) and a high conservation of a-a composition allowed by very low substitution rate and/or negative selection pressure on nucleotide sequences. Among the four selected sORFs, two were identified as encoding for the putative orthologs of human humanin and SHLP6 while the others were highly conserved at least among birds suggesting important biological roles for the putative encoded peptides. Altogether, these findings extend to birds the notion already characterised in human and rodents that sORFs embedded in mitochondrial rDNA may encode biologically active peptides.

Author contributions

MM collected data, aligned sequences, analyzed and interpreted data and prepared the first draft of the manuscript. CDu and CR conceived the study and supervised the analysis. CDe contributed to the analysis of data. All authors discussed the results and implications, edited the manuscript and gave final approval for publication.

Acknowledgements

The authors thank Christophe Douady, Tristan Lefebure and Laurent Duret for helpful discussions and suggestions. The project was financially supported by the Centre National

de la Recherche Scientifique and the University of Lyon. MM was in receipt of a fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche.

Conflict of interest

The authors report no conflict of interest.

References

[1] Papa S, Martino PL, Capitanio G, Gaballo A, De Rasmo D, Signorile A, Petruzzella V. The oxidative phosphorylation system in mammalian mitochondria. *Adv Exp Med* 942: 3-37, 2012. doi:10.1007/978-94-007-2869-1_1.

[2] Wallace DC, Fan W, Procaccio V. Mitochondrial energetics and therapeutics. *Annu Rev Pathol* 5: 297-348, 2010. doi:10.1146/annurev.pathol.4.110807.092314.

[3] Shokolenko IN, Alexeyev MF. Mitochondrial DNA: A disposable genome? *Biochim Biophys Acta* 1852: 1805-1809, 2015. doi:10.1016/j.bbadis.2015.05.016.

[4] Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res* 27: 1767-1780, 1999. doi:10.1093/nar/27.8.1767.

[5] Breton S, Milani L, Ghiselli F, Guerra D, Stewart DT, Passamonti M. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet* 30: 555-564, 2014. doi:10.1016/j.tig.2014.09.002.

[6] Hashimoto Y, Niikura T, Tajima H, Yasukawa T, Sudo H, Ito Y, Kita Y, Kawasumi M, Kouyama K, Doyu M, Sobue G, Koide T, Tsuji S, Lang J, Kurokawa K, Nishimoto I. A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. *Proc Natl Acad Sci USA* 98: 6336-6341, 2001a. Erratum in: *Proc Natl Acad Sci USA* 98, 2001. doi: 10.1073/pnas.101133498.

[7] Hashimoto Y, Ito Y, Niikura T, Shao Z, Hata M, Oyama F, Nishimoto I. Mechanisms of neuroprotection by a novel rescue factor humanin from Swedish mutant amyloid precursor protein. *Biochem Biophys Res Commun* 283: 460-468, 2001b. doi:10.1006/bbrc.2001.4765.

[8] Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R, Cohen P. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* 21: 433-454, 2015. doi:10.1016/j.cmet.2015.02.009.

[9] Cobb LJ, Lee C, Xiao J, Yen K, Wong RG, Nakamura HK, Mehta HH, Gao Q, Ashur C, Huffman DM, Wan J, Muzumdar R, Barzilai N, Cohen P. Naturally occurring

488 mitochondrial-derived peptides are age-dependent regulators of apoptosis, insulin
489 sensitivity, and inflammatory markers. *Aging* 8: 786-809, 2016.
490 doi:10.18632/aging.100943.
491 [10] Lee C, Yen K, Cohen P. Humanin: a harbinger of mitochondrial-derived peptides?
492 *Trends Endocrinol Metab* 24: 222-228, 2013. doi:10.1016/j.tem.2013.01.005.
493 [11] Kim SJ, Xiao J, Wan J, Cohen P, Yen K. Mitochondrially derived peptides as novel
494 regulators of metabolism. *J Physiol* 595: 6613-6621, 2017. doi:10.1113/JP274472.
495 [12] Capt C, Passamonti M, Breton S. The human mitochondrial genome may code for
496 more than 13 proteins. *Mitochondrial DNA A DNA Mapp Seq Anal* 27: 3098-3101,
497 2016. doi:10.3109/19401736.2014.1003924.
498 [13] Logan IS. Pseudogenization of the Humanin gene is common in the mitochondrial DNA
499 of many vertebrates. *Zoological Res* 38:198-202, 2017. doi: 10.24272/j.issn.2095-
500 8137.2017.049
501 [14] Ladoukakis E, Pereira V, Magny EG, Eyre-Walke, A, Couso JP. Hundreds of putatively
502 functional small open reading frames in Drosophila. *Genome Biol* 12: R118, 2011.
503 doi:10.1186/gb-2011-12-11-r118.
504 [15] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
505 throughput. *Nucleic Acids Res* 32: 1792-1797, 2004. doi :10.1093/nar/gkh340.
506 [16] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis
507 Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33: 1870-1874, 2016.
508 doi:10.1093/molbev/msw054.
509 [17] Vingron M, Brazma A, Coulson R, Van Helden J, Manke T, Palin K, Sand O, Ukkonen
510 E. Integrating sequence, evolution and functional genomics in regulatory genomics.
511 *Genome Biology* 10: 202-209, 2009. doi:10.1186/gb-2009-10-1-202.
512 [18] Rozas J. DNA sequence polymorphism analysis using DnaSP. *Methods Mol Biol* 537:
513 337-350, 2009. doi:10.1007/978-1-59745-251-9_17.
514 [19] Jukes TH, Cantor CR. *Evolution of protein molecules.* In: *Mammalian Protein*
515 *Metabolism*, edited by Munro HN, Academic Press, New York, 1969, p.21-132.
516 doi:10.1016/B978-1-4832-3211-9.50009-7.
517 [20] Lynch M, Crease TJ. The analysis of population survey data on DNA sequence
518 variation. *Mol Biol Evol* 7: 377-394, 1990. doi:10.1093/oxfordjournals.molbev.a040607.
519 [21] Nei M. *Molecular Evolutionary Genetics. New York: Columbia University Press*, 1987.
520 doi:10.1002/ajpa.1330750317.
521 [22] Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and
522 nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418-426, 1986.
523 doi:10.1093/oxfordjournals.molbev.a040410.

[23] Nei M, Miller JC. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* 125: 873-879, 1990.

[24] Osawa S, Jukes TH, Watanabe K, Muto A. Recent evidence for evolution of the genetic code. *Microbiol Rev* 56: 229-264, 1992. doi:0146-0749/92/010229-36$02.00/0.

[25] Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Pop Biol* 7: 256-276, 1975. doi:10.1016/0040-5809(75)90020-9.

[26] Roussel D, Boël M, Mortz M, Romestaing C, Duchamp C, Voituron Y. Threshold effect in the H_2O_2 production of skeletal muscle mitochondria during fasting and refeeding. *J Exp Biol* 222, jeb196188, 2019. doi:10.1242/jeb.196188.

[27] Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55: 541-555, 2003. doi:10.1016/j.mimet.2003.08.009.

[28] Wang Y, Tian RM, Gao ZM, Bougouffa S, Qian PY. Optimal eukaryotic 18S and universal 16S/18S ribosomal RNA primers and their application in a study of symbiosis. *PLoS One* 9, 2014. doi:10.1371/journal.pone.0090053.

[29] Gutell RR, Larsen N, & Woese CR. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiological Reviews*, *58*(1), 10–26, 1994.

[30] Fickett JW. ORFs and genes: how strong a connection? *J Comput Biol* 2: 117-123, 1995. doi:10.1089/cmb.1995.2.117.

[31] Guigó R, Fickett JW. Distinctive sequence features in protein coding genic non-coding, and intergenic human DNA. *J Mol Biol* 253: 51-60, 1995. doi:10.1006/jmbi.1995.0535.

[32] Caricasole A, Bruno V, Cappuccio I, Melchiorri D, Copani A, Nicoletti F. A novel rat gene encoding a Humanin-like peptide endowed with broad neuroprotective activity. *FASEB J* 16: 1331-1333, 2002. doi:10.1096/fj.02-0018fje.

[33] Kariya S, Hirano M, Furiya Y, Ueno S. Effect of humanin on decreased ATP levels of human lymphocytes harboring A3243G mutant mitochondrial DNA. *Neuropeptides* 39: 97-101, 2005. doi:10.1016/j.npep.2004.11.004.

[34] Thummasorn S, Shinlapawittayatorn K, Khamseekaew J, Jaiwongkam T, Chattipakorn SC, Chattipakorn N. Humanin directly protects cardiac mitochondria against dysfunction initiated by oxidative stress by decreasing complex I activity. *Mitochondrion* 38: 31-40, 2018. doi:10.1016/j.mito.2017.08.001.

[35] Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* 15, 1987. doi:10.1093/nar/15.20.8125.

[36] Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299(1–2), 1–34, 2002. doi:10.1016/S0378-1119(02)01056-9.

[37] Noderer WL, Flockhart RJ, Bhaduri A, Diaz de Arce AJ, Zhang J, Khavari PA, & Wang CL. Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Molecular Systems Biology* 10(8), 748, 2014. doi:10.15252/msb.20145136.

[38] Park J, Kim J, Park W. Environmental stress to Ethiopia indigenous chicken breeds induce immune signaling response among transcriptomic changes in heart, breast muscle and spleen tissues. Public on Sep. 13, 2018. Available in GEO DataSets (NCBI), ID of accession: GSE119387.

[39] Wang J. Transcriptomic Analysis between Normal and High-intake Feeding Geese Provides Insight into Adipose Deposition and Susceptibility to Fatty Liver in Migratory Birds. Public on Jan. 01, 2019. Available in GEO DataSets (NCBI), ID of accession: GSE119421.

[40] Nei M, Kumar S. *Molecular Evolution and Phylogenetics*, Oxford University Press, 2000. doi:10.1046/j.1365-2540.2001.0923a.x.

[41] Krejcova G, Patocka J, & Slaninova J. Effect of humanin analogues on experimentally induced impairment of spatial memory in rats. *Journal of Peptide Science*, 10(10), 636–639, 2004. doi:org/10.1002/psc.569.

[42] Hashimoto Y, Niikura T, Ito Y, Sudo H, Hata M, Arakawa E, Nishimoto I. Detailed characterization of neuroprotection by a rescue factor humanin against various Alzheimer's disease-relevant insults. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 21(23), 9235–45, 2001c. doi:10.1523/JNEUROSCI.21-23-09235.2001.

[43] Wang D, Zhang S, He F, Zhu J, Hu S, Yu J. How do variable substitution rates influence Ka and Ks calculations? *Genomics Proteomics Bioinformatics* 7: 116-27, 2009. doi:10.1016/S1672-0229(08)60040-6.

[44] Wang D, Liu F, Wang L, Huang S, Yu J. Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biology Direct* 6-13, 2011. doi:10.1186/1745-6150-6-13.

[45] Xing Y, Lee C. Can RNA selection pressure distort the measurement of Ka/Ks? *Gene* 370: 1-5, 2006. doi:10.1016/j.gene.2005.12.015.

[46] Alekhina OM, & Vassilenko KS. Translation initiation in eukaryotes: Versatility of the scanning model. *Biochemistry (Moscow)*, *77*(13), 1465–1477, 2012. http://doi.org/10.1134/S0006297912130056.

[47] Kochetov AV. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *BioEssays*, *30*(7), 683–691. doi:10.1002/bies.20771.

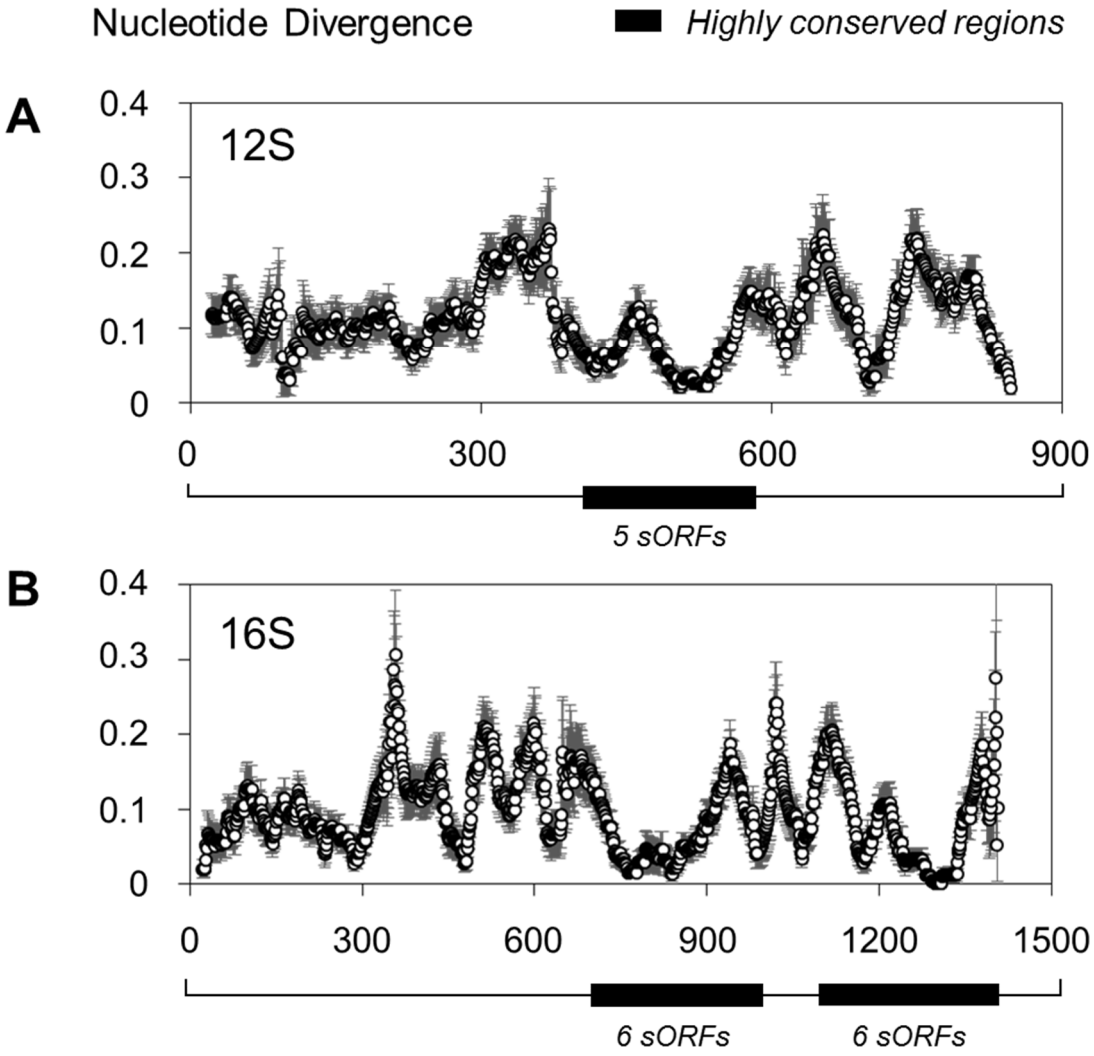


Fig. 1. Nucleotide divergences along 12S (**A**) and 16s (**B**) avian rDNAs (552 avian species). Divergences were calculated for all positions without InDel events from the alignment of 12S or 16S rDNAs. Values are the mean divergences ± SEM of a 45 nucleotides sliding window. Highly conserved regions identified in 12S or 16S rDNA (p-value < 10^{-6}) are represented by dark rectangles. Note that other regions were significantly conserved ($p<0.05$) (12S: positions 213-300; 16S: positions 51-155 and 223-317). The consensus sequences of highly conserved regions were used to detect highly conserved sORFs (≥ 45 nucleotides) in 12S (n=5) and 16S (n=11) rDNAs by referring to mitochondrial and/or standard genetic codes.

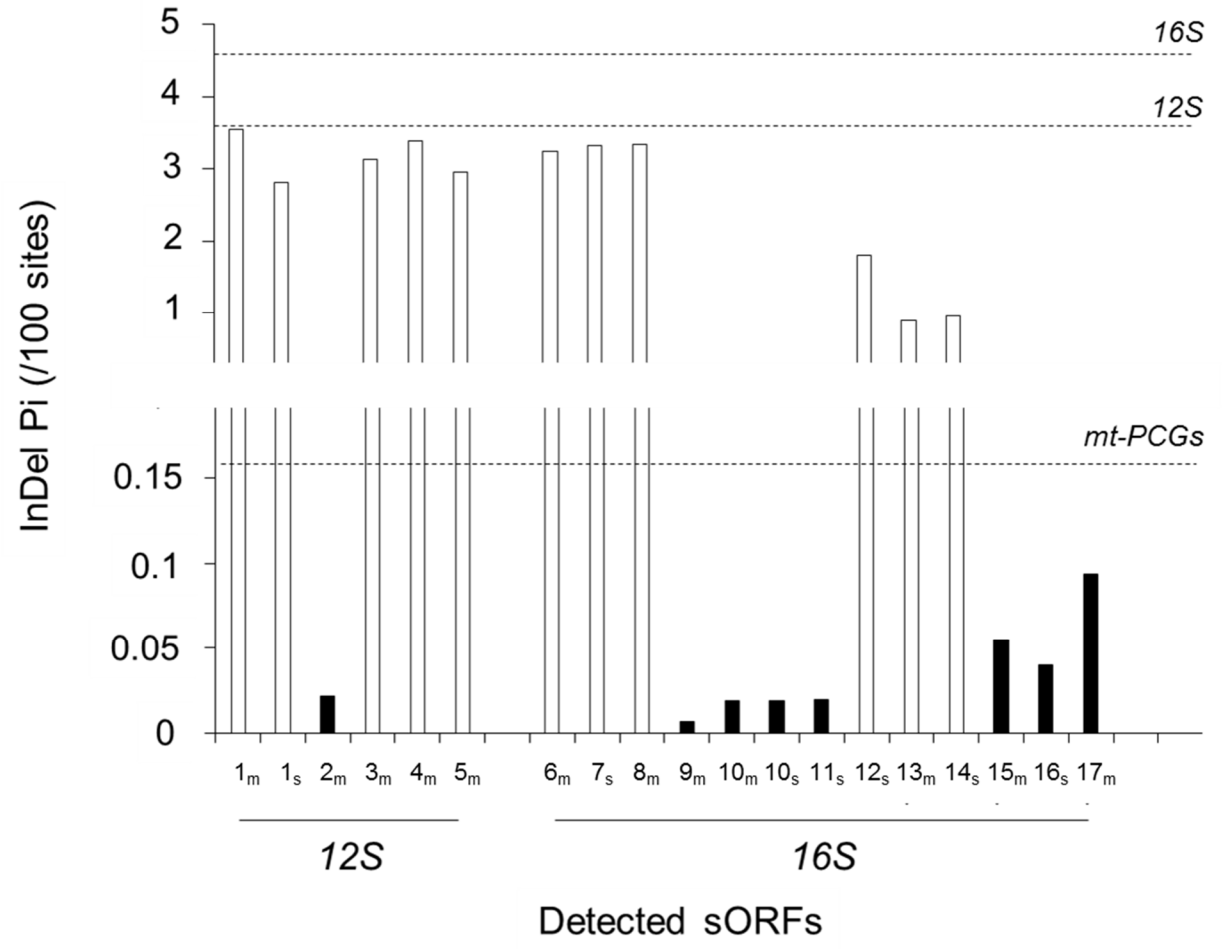


Fig. 2. InDel polymorphism rates (InDel Pi) of detected sORFs in 12S or 16S rDNAs referring to either mitochondrial (m) or standard (s) genetic code. InDel Pi values calculated for the entire 12S or 16S rDNAs and for the concatenate of the thirteen mt-PCGs are indicated by dotted lines. White bars represent sORFs with InDel Pi values close to those measured for the entire 12S (3.68%) or 16S (4.6%) rDNAs. Dark bars represent sORFs with InDel Pi values lower than that calculated for the concatenate of the 13 mt-PCGs (0.16%). Note that the y axis was broken and two scales are used for clarity.

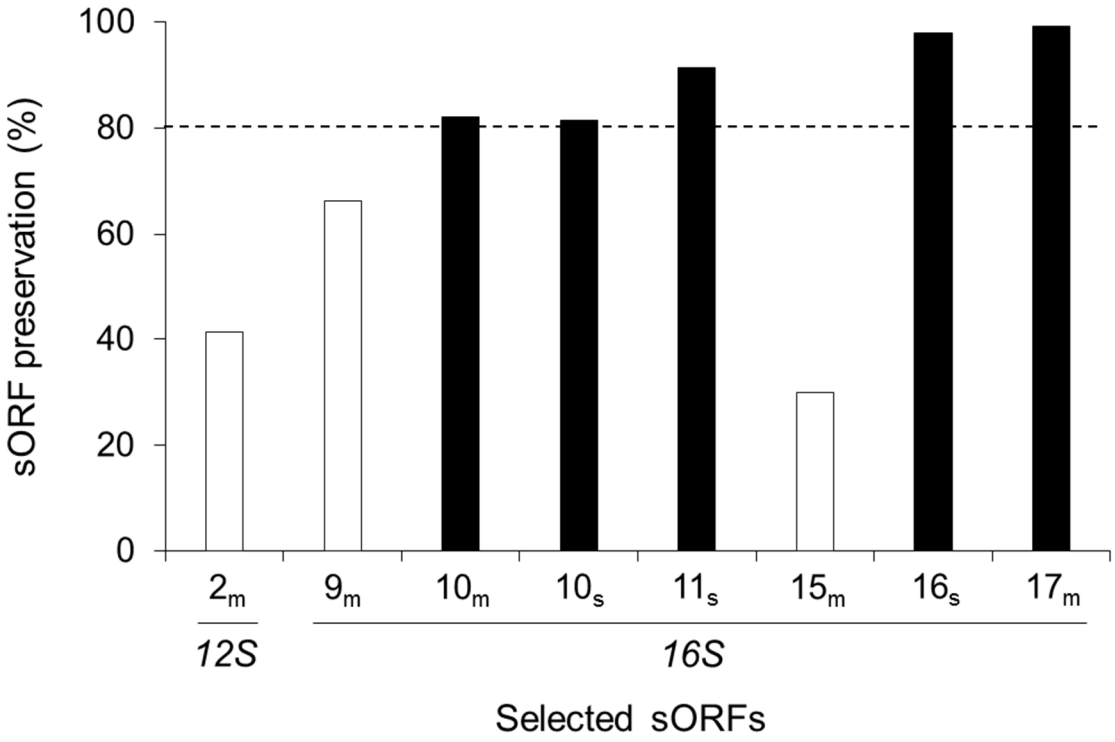


Fig. 3. sORF preservation of seven selected sORFs in bird species. The percentage of preservation corresponds to the proportion of bird species presenting start and stop codons in frame with initially detected consensus sORFs and flanking a region of ≥45 nucleotides. Dark histograms represent the four sORFs preserved in more than 80% of bird species. Note that sORF 9_{ms} is detected by using both mitochondrial and standard genetic codes.

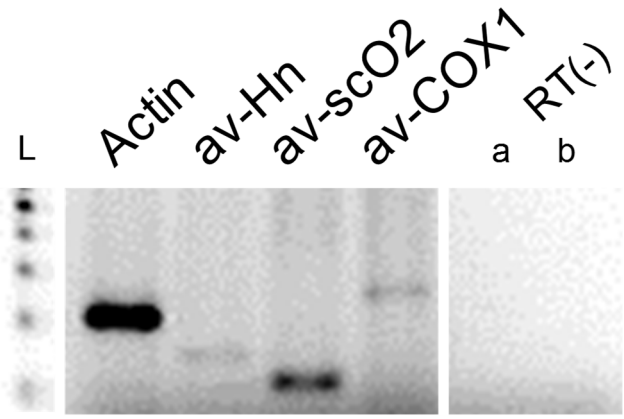
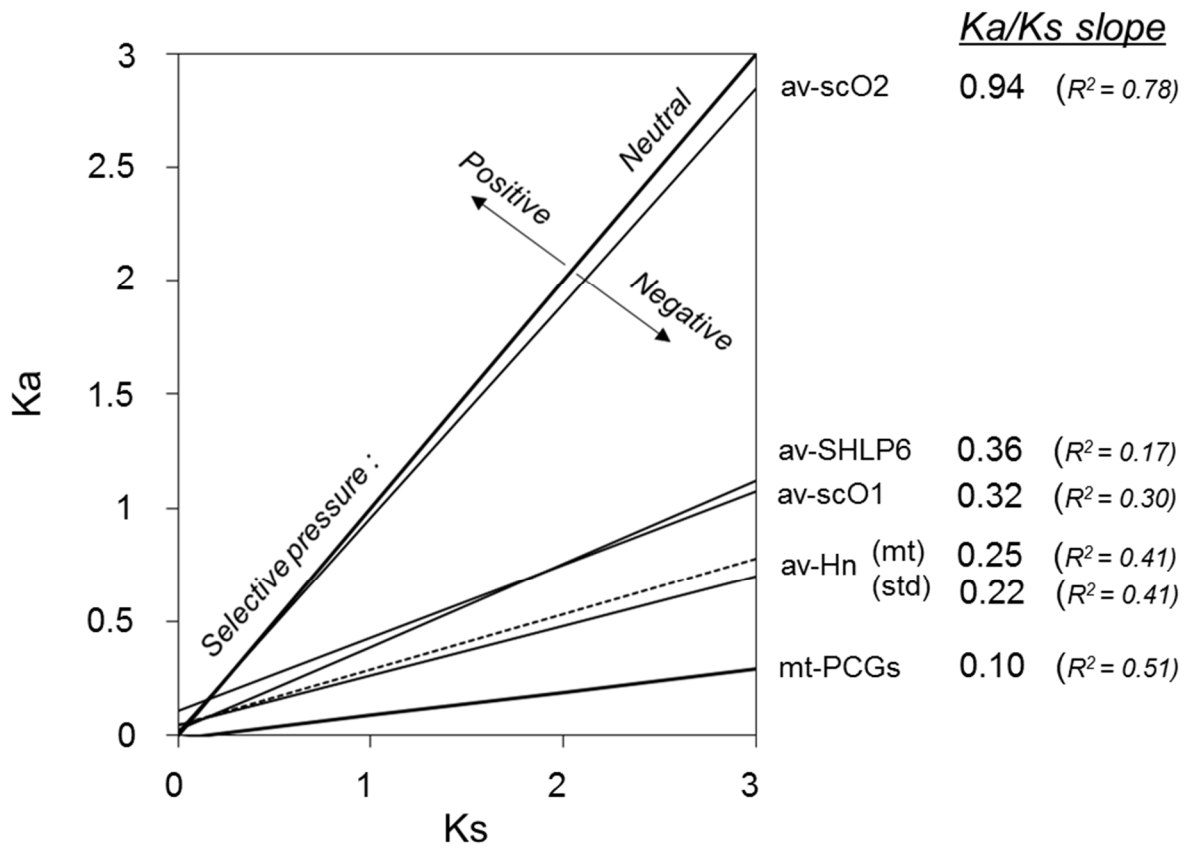


Fig. 4. RT-PCR analysis of the expression of av-Hn and av-scO2 in pectoralis muscle from a 6-wk-old muscovy duckling. RT-PCR reactions were performed using mRNAs that were purified from total RNA extracted from muscle tissue with Dynabeads® mRNA purification kit after DNAse I digestion to remove traces of DNA and 30 cycles of PCR amplification were used. Expected sizes were 155 bp for av-Hn and 127 bp for av-scO2. As positive controls, avian actin (225 bp), encoded by a nuclear gene, and avian cytochrome c oxidase subunit 1 (av-Cox1, 230 bp), encoded by a mitochondrial gene, were used. As negative controls, PCR amplifications of RNA not submitted to reverse transcription (RT(-)) were used with primers for av-Hn (a) or av-scO2 (b) and poly(dT) to verify that only polyadenylated transcripts were detected with RT-PCR. A 100 bp ladder (L) is shown.



658

659

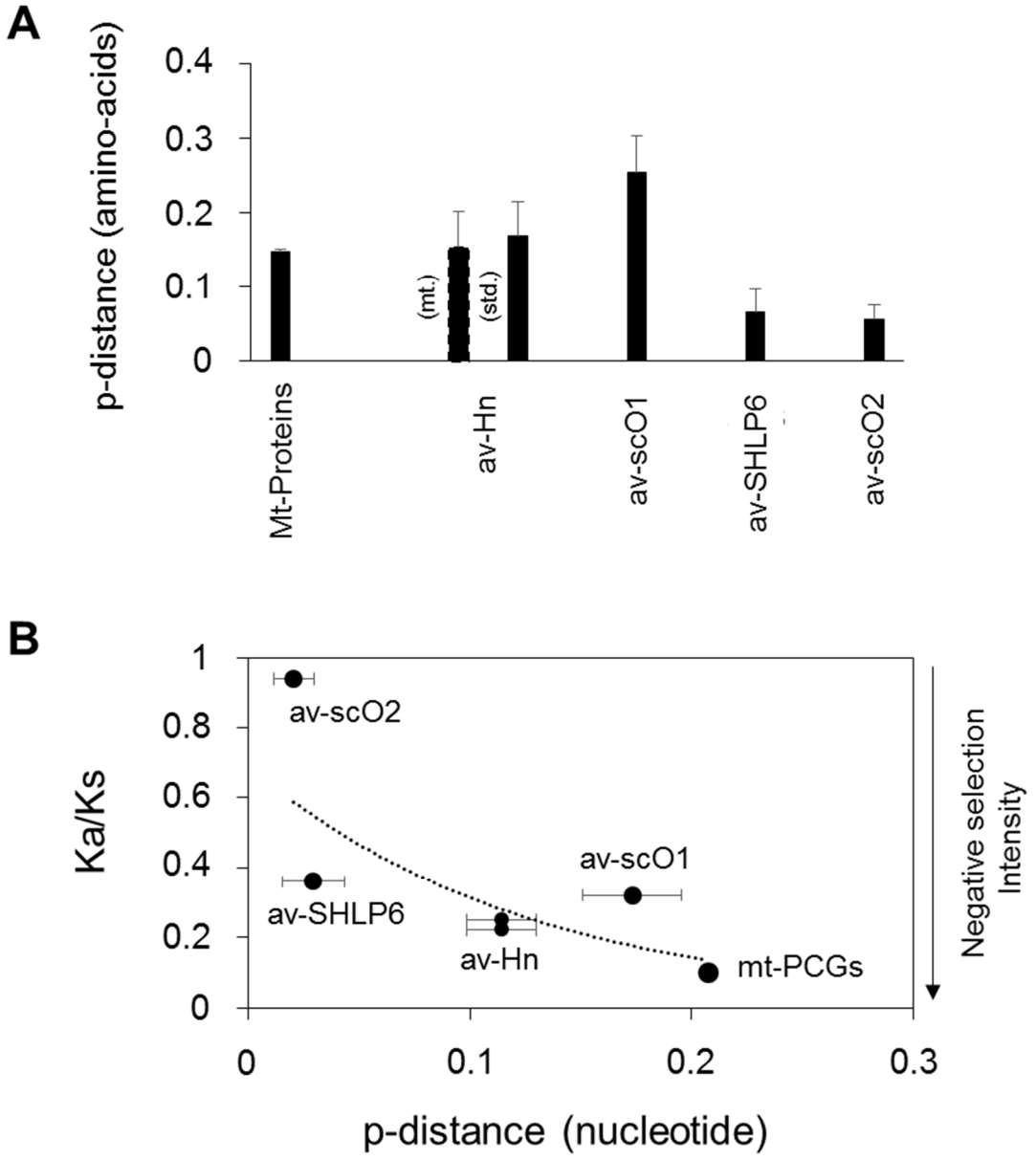
660 **Fig. 5.** Relation between Ka and Ks for the selected avian sORFs and avian mitochondrial
661 protein coding genes (mt-PCGs). Trend lines were drawn for each sORF and for the
662 concatenate of avian mt-PCGs and their respective slope (Ka/Ks) and R squared (R^2) values
663 are indicated. These plots were obtained for each DNA region by calculating Ka and Ks
664 values for pairwise comparisons among all studied bird species (552 bird species, 152 352
665 pairwise comparisons per plot). The slope of trend lines is a good indicator of the selective
666 pressure imposed on the nucleotide region, by indicating either positive (Ka>Ks), neutral
667 (Ka=Ks) or negative selection (Ka<Ks).

668

669

670

671



672

673 **Fig. 6.** Conservation of a-a sequence for the proteins / peptides of interest (**A**) and relation
674 between the rate of substitution and the negative selection intensity in birds. Proteins /
675 peptides of interest were those potentially encoded by the selected sORFs and the
676 concatenate of the known proteins encoded by mitochondrial protein-coding genes (mt-
677 PCGs). The p-distance among all 552 studied bird species was calculated by using either
678 translated peptide (**A**) or nucleotide (**B**) sequences. Variances were estimated through 100
679 bootstrap replications of each calculation. The relation between the rate of substitution and
680 the negative selection intensity (**B**) was drawn by plotting Ka/Ks ratio as a function of
681 nucleotide p-distance value for each region. The smaller the Ka/Ks ratio, the higher the
682 relative intensity of negative selection.

Table 1

Comparison of avian selected sORFs with potential human counterparts.

sORF	Position	Gen. Code	Strand	Species	a-a sequence	Identification
10_{ms}	886-960	Mt/Std	+	Birds	**MA**K**RG**LN**CLL**QVI***SEID***LPV**QKQG*	av-Hn (*avian orthologue of Humanin*)
				Human	**MA**P**RG**FS**CLL**LLT***SEID***LPV**KRRA*	Humanin (*Hashimoto et al, 2001*)
16_{s}	1238-1300	Std	+	Birds	**MLDQDI**L**MVQPLL**R**VRLFND***	av-SHLP6 (*avian orthologue of SHLP6*)
				Human	**MLDQDI**P**MVQPLL**K**VRLFND***	SHLP6 (*Cobb et al , 2016*)
11_{s}	913-966	Std	-	Birds	MCVFLL**L**YREI**NFT**DYL*	av-scO1 (*avian strongly conserved ORF1*)
				Human	*AVLCPP**L**HGQV**NFT**G*K**	No homologous sORF
17_{m}	1274-1351	Mt	-	Birds	**ME*TDLD***C**SGLNSDHVGL**LIVEQTNP*	av-scO2 (*avian strongly conserved ORF2*)
				Human	**ME*TDLD***Y**SGLNSDHVGL***	Homologous sORF not yet identified

Bird consensus sequences and their human (NC_012920.1, Pos.1671:3229) homologous regions were aligned by codons and translated into a-a sequences using either mitochondrial and/or standard genetic (gen.) codes. Putative orthologs of human peptides were renamed accordingly. sORFs with no human counterpart so far were also renamed. The position of each sORF in consensus sequence and putative encoding strand are specified. Bold letters indicate conserved a-a in birds and human. Analysis of putative peptides with the PROSITE database (https://prosite.expasy.org/) led to identify functional domains in av-Hn and av-scO2 sequences such as putative casein kinase II phosphorylation sites (underlined bold italic letters) that are fully conserved between birds and human in either av-Hn or av-scO2.