

3-21-2022

## Mechanisms of mitochondrial promoter recognition in humans and other mammalian species


Angelica Zamudio-Ochoa  
*Thomas Jefferson University*

Yaroslav I Morozov  
*Thomas Jefferson University*

Azadeh Sarfallah  
*Thomas Jefferson University*

Michael Anikin  
*Rowan University*

Dmitry Temiakov  
*Thomas Jefferson University*  
Follow this and additional works at: <https://jdc.jefferson.edu/bmpfp>

 Part of the [Medical Biochemistry Commons](#), and the [Medical Molecular Biology Commons](#)

[Let us know how access to this document benefits you](#)

### Recommended Citation

Zamudio-Ochoa, Angelica; Morozov, Yaroslav I; Sarfallah, Azadeh; Anikin, Michael; and Temiakov, Dmitry, "Mechanisms of mitochondrial promoter recognition in humans and other mammalian species" (2022). *Department of Biochemistry and Molecular Biology Faculty Papers*. Paper 212.  
<https://jdc.jefferson.edu/bmpfp/212>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Biochemistry and Molecular Biology Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

# Mechanisms of mitochondrial promoter recognition in humans and other mammalian species

Angelica Zamudio-Ochoa<sup>1</sup>, Yaroslav I. Morozov<sup>1</sup>, Azadeh Sarfallah<sup>1</sup>, Michael Anikin<sup>2</sup> and Dmitry Temiakov<sup>1,\*</sup>

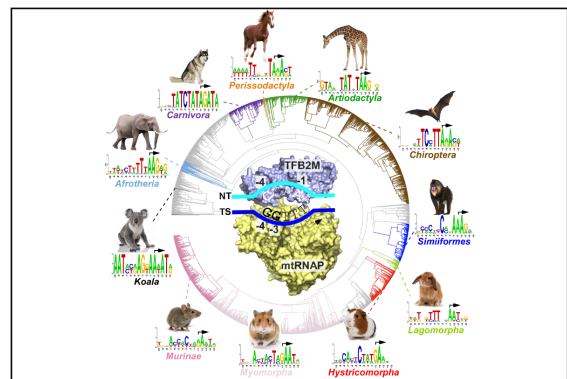
<sup>1</sup>Department of Biochemistry and Molecular Biology, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107, USA and <sup>2</sup>Department of Cell Biology and Neuroscience, Rowan University, School of Osteopathic Medicine, 42 E Laurel Rd, Stratford, NJ 08084, USA

Received December 27, 2021; Revised January 27, 2022; Editorial Decision January 28, 2022; Accepted February 03, 2022

## ABSTRACT

Recognition of mammalian mitochondrial promoters requires the concerted action of mitochondrial RNA polymerase (mtRNAP) and transcription initiation factors TFAM and TFB2M. In this work, we found that transcript slippage results in heterogeneity of the human mitochondrial transcripts *in vivo* and *in vitro*. This allowed us to correctly interpret the RNAseq data, identify the *bona fide* transcription start sites (TSS), and assign mitochondrial promoters for > 50% of mammalian species and some other vertebrates. The divergent structure of the mammalian promoters reveals previously unappreciated aspects of mtDNA evolution. The correct assignment of TSS also enabled us to establish the precise register of the DNA in the initiation complex and permitted investigation of the sequence-specific protein-DNA interactions. We determined the molecular basis of promoter recognition by mtRNAP and TFB2M, which cooperatively recognize bases near TSS in a species-specific manner. Our findings reveal a role of mitochondrial transcription machinery in mitonuclear coevolution and speciation.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Mitochondrial gene expression involves interplay between nuclear-encoded mitochondrial proteins and the mitochondrial genome. The rates of mutation in mitochondria are significantly higher than in the nucleus, driving a constant adaptation process, known as mitonuclear coevolution, and facilitating speciation events (1). The control region of mitochondrial DNA, which harbors two promoters, LSP and HSP, is the most divergent region in mammalian species. The sequence and organization of the mitochondrial promoters bear no similarity to the bacterial or phage promoters despite the endosymbiotic origin of mitochondria and the relation of mtRNAP to the family of T7-like RNA polymerases (2). Hence, the mitochondrial promoters are challenging to predict and identify despite the conserved arrangement of the mitochondrial genes in mammalian species. The lack of this information has impeded our understanding of promoter recognition and speciation mechanisms driven by mitonuclear incompatibility.

Over the past decade, our knowledge of the transcription initiation process in mitochondria has improved due to a cohort of biochemical and structural studies (3–9). The sequential model of initiation postulates that mtRNAP is

\*To whom correspondence should be addressed. Email: dmitry.temiakov@jefferson.edu

recruited to the promoter by transcription factor TFAM, which is bound to the -39 to -17 region upstream of the transcription start site, TSS (6,10). The bending of the DNA by TFAM allows the establishment of specific interactions between its C-terminal ‘tail’ and the long ‘tether’ helix in the N-terminal domain of mtRNAP (3). The resulting pre-initiation complex (preIC) is transcription-incompetent and requires another initiation factor, TFB2M, to melt the promoter and initiate transcription (10). The more upstream region of the bent DNA (-40 to -60) is located in proximity to mtRNAP and appears to contribute to the overall stability of the IC; however, sequence-specific determinants have not been identified in this region (10). A conserved structural element in all single-subunit T7-like RNAPs, the specificity (SP) loop, has been proposed to interact with the -10 to -5 region of the yeast mitochondrial promoter (11); however, structural studies have not revealed such interactions in the mammalian system (3). Previous studies suggested that mutations that affected the human mitochondrial transcription initiation involved bases -1/-2 and -3/-4 in the LSP promoter (7). Whether TFB2M is involved in specific interaction with this region and thus would be a functional analog of the bacterial specificity factor  $\sigma$  had been a subject of discussion; however, no evidence of its involvement in base recognition has been presented (12).

Earlier attempts to understand the mechanisms of species-specific promoter recognition in mitochondria employed murine promoters, the only other well-established *in vitro* transcription system for mammalian species (13). The mouse and human LSPs share superficial homology, complicating the interpretation of the mutagenesis data. Interestingly, it has been determined that mouse TFAM can activate transcription on LSP when used with human mtRNAP and TFB2M (7). However, both human mtRNAP and TFB2M showed strict specificity towards human mitochondrial promoters, defined by the interactions mapped to the vicinity of the TSS (7).

The nuclear-encoded proteins that comprise the mitochondrial transcription machinery share 70–90% similarity in mammalian species. Considering an apparent lack of conservation in the promoter region of mtDNA in different species, it remains puzzling how proteins with such similar amino acid compositions can transcribe such divergent promoters. Knowing how the nuclear genes encoding the transcription machinery adapt to the mutations in mtDNA could lead to a greater understanding of mitonuclear coevolution and speciation.

In this study, we provide a detailed account of interactions of human mtRNAP with its promoters and define the major specificity determinants of the transcription initiation complex. Our data implicate transcription initiation factor TFB2M in sequence-specific interactions with the promoter and demonstrate that substitution of only three bases in the human mitochondrial promoter allows for its recognition by porcine mitochondrial transcription machinery. We also identified promoter regions and transcription start sites for most orders of mammalian species and some other vertebrates, enabling future analysis of the evolution of sequence-specific DNA recognition in mitochondria.

## MATERIALS AND METHODS

### Expression and purification of the components of human mitochondrial transcription

Cys-less  $\Delta 42$  TFAM and  $\Delta 20$  TFB2M were expressed as previously described (3).  $\Delta 119$  mtRNAP was expressed and purified as described in (10). MtRNAP mutants were obtained by site-directed mutagenesis (QuikChange, Agilent) using the  $\Delta 119$  mtRNAP in pProEx-HTb background.

### Expression and purification of the components of porcine mitochondrial transcription

A sequence encoding the porcine mtRNAP lacking the first 107 residues ( $\Delta 107$  N-his *S.s.* mtRNAP) was amplified by PCR from porcine heart cells cDNA (Zyagen, PD-801) and cloned into the pProEx-HTb vector (Invitrogen) using the *NcoI/XhoI* restriction sites. The protein was expressed and purified as previously described for human mtRNAP using Ni-agarose affinity chromatography, heparin sepharose, and gel filtration (14).

Synthetic genes encoding porcine  $\Delta 42$  TFAM (carrying an N-terminal His<sub>6</sub> tag) and  $\Delta 20$  TFB2M (carrying a C-terminal His<sub>6</sub> tag) were cloned into the pET22b vector. The internal loop (residues 265 to 287) in the porcine TFB2M was replaced by a short GSSG linker to improve the protein solubility. Both proteins were expressed in BL21 (DE3)-RIPL cells. Porcine TFAM was induced for 2 hours at 37°C with 0.4 mM IPTG and purified using the protocol for human TFAM (Hillen, 2017). Porcine TFB2M was induced for 18 hours at 12°C with 0.1 mM IPTG and purified by affinity chromatography using a HisTrap HP column (GE Healthcare), followed by affinity chromatography using a HiTrap heparin HP column (GE Healthcare). The heparin column was equilibrated in buffer A (40 mM Tris-HCl, pH 7.9, 300 mM NaCl, 5% Glycerol, 5 mM  $\beta$ -mercaptoethanol) and porcine TFB2M was eluted by 0–80% linear gradient of buffer B (40 mM Tris-HCl, pH 7.9, 1.5 M NaCl, 5% Glycerol, 5 mM  $\beta$ -mercaptoethanol). Peak fractions were pooled, concentrated, and subjected to cation exchange chromatography using Mono S 5/50 GL column (GE Healthcare). The Mono S column was equilibrated in buffer A, and porcine TFB2M was eluted by 0–80% linear gradient of buffer B. Peak fractions were pooled, concentrated, and stored at -80 °C.

### Preparation of templates for transcription assays

The porcine regulatory region (nucleotides 1 to 1175 of the *Sus scrofa* genome, NC\_000845) was cloned into the pT7Blue vector using the pT7Blue Perfectly Blunt cloning kit (Novagen). For the porcine LSP template preparation, the region including nucleotides 1011–1139 (-108 to + 21 of *Sus scrofa* LSP) was amplified by PCR and purified using a PCR purification kit (Thermo). For porcine HSP preparation, the region including nucleotides 1064–1175 (-79 to + 33 of *Sus scrofa* HSP) was amplified. Templates having single base-pair substitutions in LSP were generated using reverse PCR primers with corresponding mutations.

Synthetic templates (IDT DNA, Supplementary Table S1) containing the -41 to +19 region of human LSP or -63 to +19 region of human HSP were used for comprehensive mutagenesis. For annealing, the DNA oligonucleotides (20  $\mu$ l) were diluted in the 'duplex buffer' (100 mM potassium acetate; 30 mM HEPES, pH 7.5) to 5  $\mu$ M concentration, heated for 7 min at 95 °C and cooled down (1 °C/min) for 70 min to 25 °C in a thermocycler. Upon annealing, the templates were diluted to 0.5  $\mu$ M concentration with water.

The human templates for the heterologous experiments and the mutant mtRNAP transcription assays were obtained by PCR amplification. The region comprising nucleotides 383 to 597 of the human reference genome NC.012920.1 was cloned into the pT7Blue vector using the pT7Blue Perfectly Blunt cloning kit (Novagen, 70189). For human LSP preparation, nucleotides 386 to 514 (from -108 to +21 of human LSP) were amplified. For human HSP preparation, nucleotides 456 to 586 (-106 to +25 of human HSP) were amplified. Templates having single base-pair substitutions in LSP were generated using a reverse PCR primer with the desired mutation.

### Transcription assays

Transcription reactions were carried out using synthetic or PCR DNA templates (50 nM), mtRNAP (150 nM), TFAM (100 nM for synthetic templates or 200 nM for PCR templates) and TFB2M (150 nM) in a transcription buffer containing 50 mM Tris (pH = 7.9), 10 mM MgCl<sub>2</sub>, 20 mM  $\beta$ -mercaptoethanol and 0.1 mg/ml of BSA in the presence of ATP (0.3 mM), GTP (0.3 mM), CTP (0.3 mM), UTP (0.05 mM) and 0.3  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P] UTP (800 Ci/mmol). When the mutant mtRNAPs were used, reactions were carried out at 34 °C for 5 min. All the other transcription assays were carried out at 37 °C for 30 min. Reactions were stopped by the addition of an equal volume of 95% formamide/0.05 M EDTA. The products were resolved by 20% PAGE containing 6 M Urea and visualized by PhosphorImager (GE Health). All experiments were repeated at least three times, and the representative images are shown in the figure.

### Analysis of the catalytic activity of mtRNAP mutants by primer extension

The activity of mtRNAP and its variants was assayed using a primer extension assay (15). A 14 nt RNA (16) was labeled at its 5' end using [ $\gamma$ -<sup>32</sup>P]ATP and PNK (NEB) and was annealed to oligonucleotides TS02 and NT02 as described for synthetic templates above.

### Isolation of mitochondria

To isolate mitochondria, 10<sup>7</sup> HeLa cells were lysed using a Teflon homogenizer in 7.5 ml of mitochondria isolation buffer (20 mM HEPES-KOH, pH 7.2, 0.25 M sucrose, 1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 0.1 mM PMSF). The lysate was cleared by two cycles of centrifugation at 2,000 g for 5 minutes at 4 °C. The supernatant was collected and centrifuged at 10,000 g for 12 minutes.

The pellet containing mitochondria was washed with HES Buffer (20 mM HEPES-KOH pH 7.2, 1 mM EDTA, 0.25 M sucrose) and stored in the same buffer at -80 °C.

### Identification of the 5' ends of mitochondrial transcripts

Isolated mitochondria were taken up in 1 ml of Trizol (Thermo Fisher Scientific), incubated for 5 min at room temperature, mixed with 0.2 ml of chloroform, and vortexed for 15 s. The mixture was spun down at 12,000 g for 15 min to separate phases. The colorless upper phase was collected, mixed with 0.5 ml of isopropanol, and incubated for 15 min at room temperature. The RNA pellet was collected by centrifugation at 12,000 g for 15 min. The pellet was washed with 1 ml of 75% ethanol by centrifugation at 12,000 g for 15 min, dried, dissolved in nuclease-free water, and treated with DNase I (NEB) to remove DNA contamination. The typical yield of mitochondrial RNA was 12.5  $\mu$ g per 10<sup>6</sup> cells.

The DNA primers complementary to the initially transcribed regions of the LSP and HSP promoters (Supplementary Table S1) were <sup>32</sup>P-labeled with PNK (NEB). To anneal the primers to the RNA, the mixture containing 3–4  $\mu$ g of total RNA and DNA primer (100 nM) was heated for 5 min at 75 °C and flash-cooled on ice for 10 min. The primer extension was performed using ProtoScript II reverse transcriptase (NEB) in a reaction containing RiboLock RNase inhibitors (Thermo Fisher), 10 mM DTT, dNTPs (1 mM), and 0.75–3  $\mu$ g of total RNA for 30 min at 42 °C. Reactions were stopped by the addition of an equal volume of 95% formamide/0.05M EDTA. The products were resolved by 20% PAGE containing 6 M urea and visualized by PhosphorImager (GE Health).

### Identification of mitochondrial promoters

NCBI Sequence Read Archives (SRA) database has been used as the source of the primary data for promoter search. The dataset accession numbers are listed in Table 1. The part of the control region of mtDNA (between tRNA<sup>Phe</sup> and CSBII), which harbors mitochondrial promoters, was used as a query for BLASTn search of the datasets containing sequences of non-polyadenylated RNA. Identification of a homogeneous 5' end for multiple reads aligning to the same query region was interpreted as an indication of a transcription start site (+1). In some cases, transcripts originating from one of the mitochondrial promoters contained several additional AMP residues at their 5' ends, and therefore the exact TSS could not be identified with a single nucleotide precision. To overcome this, alignment of a putative promoter region with the identified promoter from the other mtDNA strand or with the promoters identified for closely related species has been used.

### Generation of sequence logos and phylogenetic tree

DNA sequence logos were built using the weblogo.berkeley.edu/logo.cgi server. The phylogenetic tree of extant mammalian species was generated using iTOL software (17,18).



**Table 1.** Identification of mitochondrial promoters and their TSS in mammalian species

Order	Species	RNA-seq accession number	NCBI reference sequence	LSP TSS	HSP TSS
<i>Rodentia</i>	<i>Mus musculus</i>	SRX3061644	NC_005089.1	16188	16285, 16296
<i>Rodentia</i>	<i>Rattus norvegicus</i>	SRX016597	NC_001665.2	16191	16292
<i>Rodentia</i>	<i>Ictidomys tridecemlineatus</i>	SRX2039082, SRX9106226	NC_027278.1	16285	16437
<i>Rodentia</i>	<i>Cavia porcellus</i>	SRX1527562	NC_000884.1	16741	16770
<i>Rodentia</i>	<i>Heterocephalus glaber</i>	SRX9149419	NC_015112.1	16289	16346
<i>Lagomorpha</i>	<i>Oryctolagus cuniculus</i>	SRX3028972-74, SRX3028966-67	NC_001913.1	16666*, 16819*, 16972*, 17125*	16719, 16872, 17025, 17178
<i>Lagomorpha</i>	<i>Ochotona dauurica</i>	SRX2865113, SRX2865123	NC_044120.1	16959, 17164	17034*, 17239*
<i>Primates</i>	<i>Homo sapiens</i>	SRX12687743-45	NC_012920.1	406*	562*
<i>Primates</i>	<i>Macaca mulatta</i>	SRX10270359	NC_005943.1	398*	503*
<i>Primates</i>	<i>Callithrix jacchus</i>	SRX1166092	CM021961.1	16294	16428
<i>Chiroptera</i>	<i>Phyllostomus discolor</i>	SRX6878473	CM016611.2	16606*	16625
<i>Artiodactyla</i>	<i>Ovis aries</i>	SRX5547561	NC_001941.1	16486	16588
<i>Artiodactyla</i>	<i>Sus scrofa</i>	SRX5417080	NC_000845.1	1031	1143
<i>Perissodactyla</i>	<i>Equus ferus caballus</i>	SRX9602386	NC_001640.1	16588*	16607
<i>Perissodactyla</i>	<i>Equus asinus</i>	SRX4946216	NC_001788.1	16599*	16618
<i>Carnivora</i>	<i>Canis lupus</i>	SRX3639642, SRX3639655, SRX3639682	NC_002008.4	16648	16651
<i>Carnivora</i>	<i>Felis catus</i>	SRX4999833-35	NC_001700.1	791	794
<i>Proboscidea</i>	<i>Elephas maximus</i>	SRX4745754	NC_005129.2	16811, 16800	16833
<i>Tubulidentata</i>	<i>Orycteropus afer</i>	SRX4745740	NC_002078.1	Not found	16749
<i>Cingulata</i>	<i>Dasyurus novemcinctus</i>	SRX1527560	NC_001821.1	16756*, 16819, 16947	17008*
<i>Diprotodontia</i>	<i>Phascolarctos cinereus</i>	SRX5503561 to SRX5503566	NC_008133.1	16199	16304*
<i>Monotremata</i>	<i>Ornithorhynchus anatinus</i>	SRX182799-801	NC_000891.1	Not found	16988* or 16989*
<i>Anseriformes</i>	<i>Anas platyrhynchos</i>	SRX9500716-19	NC_009684.1	910, 912	917, 919
<i>Anura</i>	<i>Xenopus laevis</i>	SRX6087549-51	NC_001573.1	2033, 2101	2104

\*Indicates transcript slippage.

## RESULTS

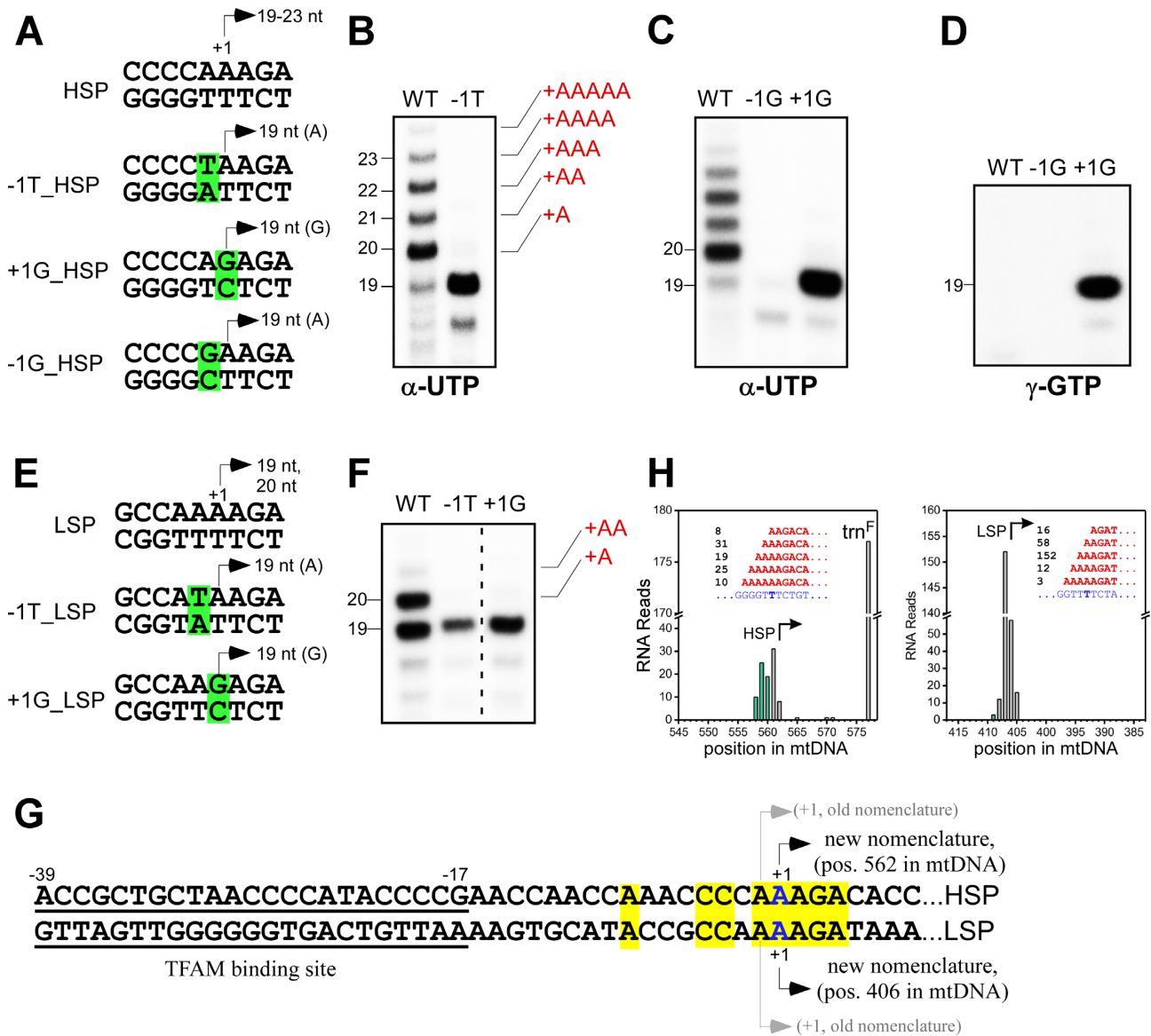
### Transcription initiation at HSP and LSP involves transcript slippage

*In vitro* transcription at human HSP generates several RNA species, some of which are 1–4 nt longer than the predicted 19–20 nt run-off product (Figure 1A, B). To understand the reason behind the heterogeneity of the HSP transcripts, we altered the base pairs in the stretch of the three template strand dTMP residues located near the TSS (Figure 1A). Strikingly, substituting the most upstream A-T base pair to a T-A base pair resulted in the synthesis of a nearly homogeneous 19 nt RNA product (Figure 1B, right lane). The observed change in transcription pattern suggests that the dTMP base in the middle of the T-stretch is the *bona fide* TSS (designated here as + 1). The additional RNA products observed during initiation at the native HSP resulted from transcription slippage, an iterative cycle of addition of AMP residues to the growing end of the transcript on a homopolymeric track of dTMP residues in the DNA (19) (Supplementary Figure S1A). To confirm this, we substituted the -1 and the + 1 A-T base pairs with G-C base pairs (Figure 1C). While transcription of the '-1G' template was inefficient, the '+1G' promoter produced a robust 19 nt long RNA product (Figure 1C). We confirmed the utilization of GTP as the priming substrate for the '+1G' promoter by using [ $\gamma$ -<sup>32</sup>P]-GTP, which can only label the 5' end of the RNA (Figure 1D, Supplementary Figure S1B). As expected, the RNA products obtained using the native HSP were not de-

tected; however, efficient labeling was observed in the case of the '+1G' HSP, confirming initiation with GTP and location of the start site at position + 1 (position 562 in the reference human mtDNA) (Figure 1D).

Transcription initiation at LSP results in two major RNA species (19 and 20 nt, Figure 1E, F), previously interpreted as originating from two adjacent alternative start sites (14). However, similar to the situation with HSP, the interruption of the T-stretch at TSS ('-1T' LSP) or mutation of the + 1 base to dGMP (and thus utilization of GTP as the priming substrate) results in the production of a single 19-nt run-off band (Figure 1F, Supplementary Figure S1C). This finding also suggests transcript slippage as the mechanism behind the formation of multiple RNA products and illuminates a single initiation site in LSP (+1) at position 406 in the human mtDNA (Figure 1G).

We next probed whether transcript slippage occurs during transcription initiation *in vivo*. We isolated total RNA from human mitochondria and performed 5'-end mapping of the HSP and LSP transcripts using primer extension by reverse transcriptase (Supplementary Figure S1D). The cDNA copies of the HSP transcripts were 1, 2, or 3 nt longer than the control <sup>32</sup>P-labeled DNA primer made to match the length and sequence of the nascent transcript originated from the + 1 position (Supplementary Figure S1D, upper panel). Similarly, in the case of the LSP transcripts, two major species were identified (Supplementary Figure S1D, lower panel). The size and the pattern of the RNA species identified at both mitochondrial promoters *in*



**Figure 1.** Transcript slippage occurs during transcription initiation at human mitochondrial promoters. (A) Schematic illustration of the mutations made near the HSP TSS. The mutated bases are highlighted in green. TSS location is indicated based on the data obtained in this work. The lengths of the observed RNA species are indicated next to the bent arrow. (B) Substitution of the -1 base-pair suppresses transcript slippage at HSP. RNA species having additional AMP residues (red) at the 5' end are indicated. (C, D) Substitution of the +1 NT base with dGMP in HSP allows transcription initiation with GTP. Transcription reactions were labeled with [ $\alpha$ - $^{32}$ P]-UTP (C) or [ $\gamma$ - $^{32}$ P]-GTP (D). (E, F) Substitution of the -1 or +1 base pairs suppresses transcript slippage at LSP. RNA species having additional AMP residues (red) at the 5' end are indicated. (G) Alignment of human HSP and LSP promoters based on the TSS identified in this work. The positions of the start sites and their location in mtDNA are indicated in blue letters and marked by black arrows. The positions of the previously identified TSS are shown in grey arrows. Identical bases between the LSP and HSP promoters in the -14 to +4 region are highlighted in yellow, and the TFAM binding site is underlined. (H) Distribution of the 5'-ends of human RNA sequence reads in the vicinity of HSP and LSP. Bars represent the number of reads in the SRA that have their 5'-ends at corresponding positions. The bars colored in teal represent the reads that have their 5'-ends mis-paired to the template DNA. Sequences of the 5'-proximal regions of the RNA species are shown in the insets (red) aligned to the sequence of the template strand of DNA (blue).

*vivo* and *in vitro* match closely, confirming the occurrence of transcript slippage during transcription initiation in human mitochondria. Finally, analysis of LSP and HSP transcripts using RNA-sequencing data (20) reveals heterogeneity at the 5' end of these transcripts consistent with the transcript slippage pattern observed in *in vitro* experiments above (Figure 1H). The observed confinement of most RNA species to only two locations within the control region of mtDNA (to LSP and HSP), and the presence of additional, mismatching AMP residues at their 5' ends, unequivocally suggests

that they represent the initially transcribed sequence and not caused by nuclease processing.

### Identification of mitochondrial promoters for mammalian species

Mammalian mitochondrial promoters are generally not easily identifiable because of the lack of sequence conservation. Identification of TSS of mitochondrial promoters is further complicated by the transcript slippage phenomenon

described in this study, resulting in the heterogeneity of the 5' ends of the transcripts, which prevents the TSS assignment with a single-nucleotide precision. Considering these findings, we analyzed the available RNAseq databases for species belonging to various orders of mammals and determined promoter regions and their TSS based on the sequence of the 5' ends of RNA transcripts (Table 1, Figure 2A, Supplementary Figures S2–S5).

Analysis of mitochondrial promoters of mammals belonging to seventeen orders reveals a different degree of conservation of the TSS region and a variable distance between LSP and HSP (Figure 2B). Overall, the TSS region is poorly conserved, with a notable exception of the order *Carnivora* (Figure 2A).

In all mtDNA sequences analyzed, LSP and HSP are found in proximity to each other (Table 1, Figure 2B). The shortest inter-promoter region is found in the *Carnivora* species (2 bp between LSP and HSP TSS), while the longest is observed in humans (155 bp). In many orders, the length of the inter-promoter region is conserved between species. Thus, the distance between the LSP and HSP TSS in all species belonging to the six orders of the superorder *Afrotheria* is 21 nt. In primates, however, the length of the inter-promoter regions differs from family to family and is between 110 and 155 nt. Surprisingly, species of the *Lagomorpha* and *Cingulata* orders have several promoter units in their mtDNA (Figure 2B). Four LSP-HSP units were detected in mtDNA of European rabbit and nine-banded armadillo, two - in Daurian pika (Figure 2B, Supplementary Figure S3).

While each species in most mammalian orders possess a distinct set of divergent promoters, promoters of some *Perissodactyla* and *Chiroptera* species share extraordinary sequence homology (Figure 2C). For example, the TSS regions in mtDNA of horses and pale spear-nosed bats are nearly identical (Figure 2C).

The *Carnivora* species have the most conserved promoter region (10–14 absolutely conserved bases in LSP/HSP) (Figures 2A, D, Supplementary Figure S4A, B). A single mutation within the palindromic region will affect both promoters and, consequently, could affect transcription and replication of mtDNA, which likely explains the high level of preservation of this region in *Carnivora* order. A closely spaced, palindromic promoter unit can be an early invention in evolution as it is also observed in amphibians (Figure 2E). Remarkably, a closely spaced AT-rich promoter unit has been identified in birds (Figure 2F). Analysis of the SRA data suggests that mtDNA of wild duck (*Anas platyrhynchos*) contains two pairs of promoters, both of which are responsible for transcription (Figure 2F, Supplementary Figure S5D). In many *Chiroptera* species, the palindromic promoter region is interrupted, and its conservation involves only the bases adjacent to the TSS of LSP and HSP (Figure 2G).

Analysis of mammalian mitochondrial promoters reveals that transcript slippage is very common and is seen in bats, rabbits, hares, horses, armadillo, platypus, and koalas on either LSP or HSP (Supplementary Figures S2–S5). The only group of animals in which the transcript slippage occurs at both promoters is the catarrhines, which includes apes and the old-world monkeys (Supplementary Figure S3).

## Comprehensive mutagenesis identifies bases essential for promoter utilization

Identifying the *bona fide* TSS in human LSP and HSP (Figure 1) allows for precise positioning of the DNA register in the active site of mtRNAP and mapping of interactions between the promoter bases and the residues in mtRNAP and transcription factors. As a consequence, probing the mechanisms behind species-specific promoter recognition is now possible.

To determine which DNA bases are critical for binding and recognition by mtRNAP, we performed comprehensive mutagenesis of the promoter region predicted to interact with mtRNAP and TFB2M (3). Human LSP is the most studied mammalian mitochondrial promoter and serves as a convenient target for mutagenesis. The bases in the region from -10 to +4 of LSP were substituted to represent all possible base-pair variants (Figure 3A, Supplementary Figure S6, and Table 2). Of the 42 promoter templates tested, the most dramatic effect on transcription efficiency was observed when the bases in proximity to the TSS were substituted (Figure 3A and Table 2). This is consistent with the conservation of these bases among human LSP and HSP in this region (Figure 1G).

As discussed above for HSP, substituting the dTMP bases that promote transcript slippage (-1A, +1A, +2A) had a notable effect on the transcription pattern (Figures 1A, B, 3A). Different substitutions in this region had variable effects on transcription efficiency, likely due to minute changes in DNA binding and melting. We also found that substituting the conserved +3G and +4A bases resulted in a 3–6-fold reduction of transcription efficiency, with a notable exception of the +3G > A mutation, which resulted in an almost complete loss of transcription initiation (Table 2).

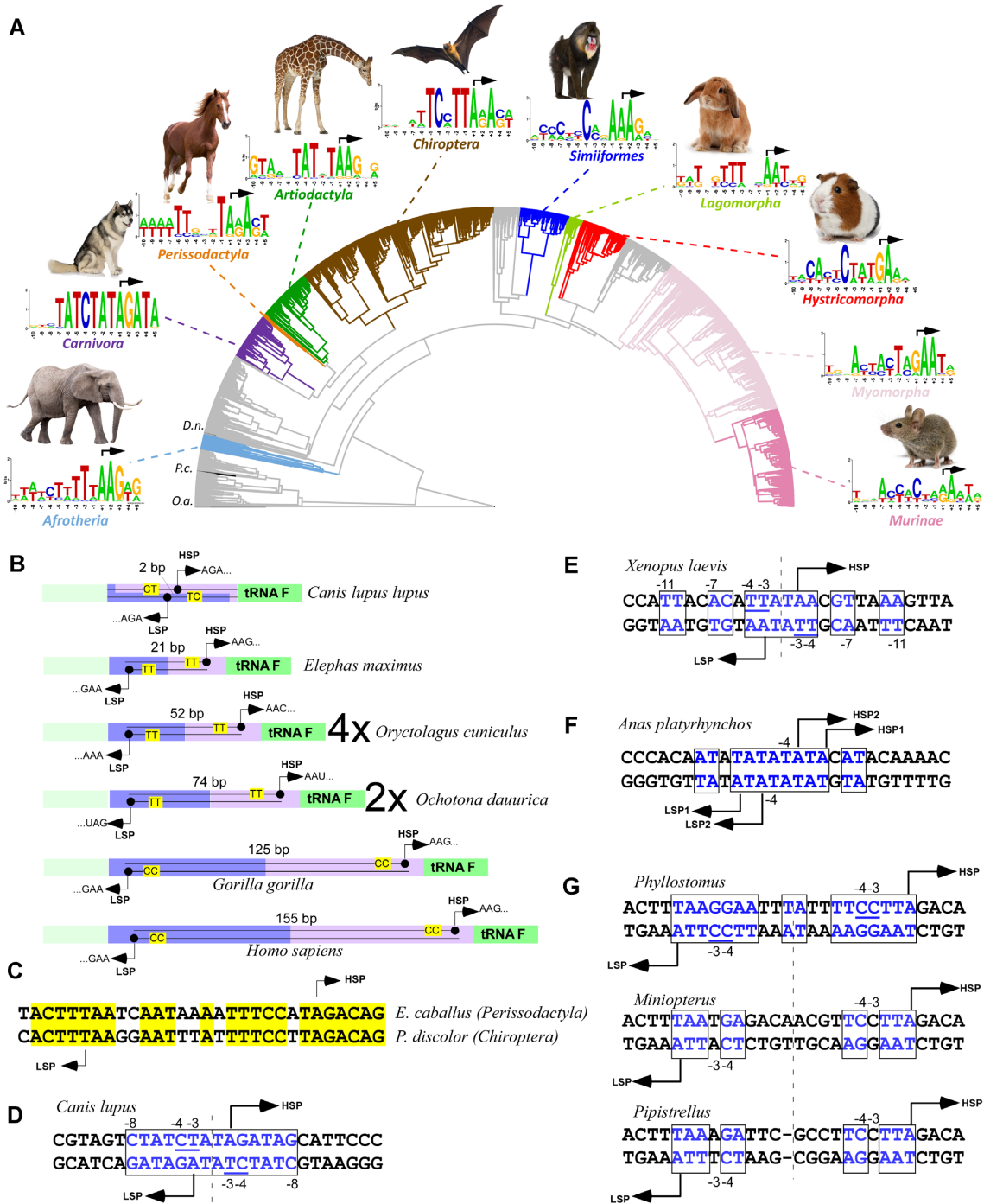
Earlier studies of mitochondrial transcription predicted that the specificity (SP) loop of yeast mtRNAP makes base-specific contacts with the promoter (11). In the human mitochondrial transcription initiation complex, the SP loop is inserted into the major groove of DNA and is thus in a position to make base-specific interactions (3). However, unlike the situation with the phage T7 promoters, the upstream region (bases from -12 to -6) is not conserved between LSP and HSP (Figure 1G), and base substitutions there do not considerably affect the efficiency of transcription initiation (Table 2). While our mutagenesis data cannot completely rule out base-specific interactions with this region of the promoter, it seems likely that the mode of promoter recognition by the SP loop in mtRNAP may differ quite significantly from that observed in phage RNAPs.

Finally, substituting the -3 and -4 base-pairs (conserved in both LSP and HSP in human mtDNA) resulted in dramatic inhibition of transcription initiation (Figure 3A, and Table 2). These bases are melted in the IC (21) and can, in theory, be recognized in either strand of DNA by mtRNAP and/or TFB2M (3).

## Promoter bases -3 and -4 are located in close proximity to conserved structural elements in mtRNAP in the IC

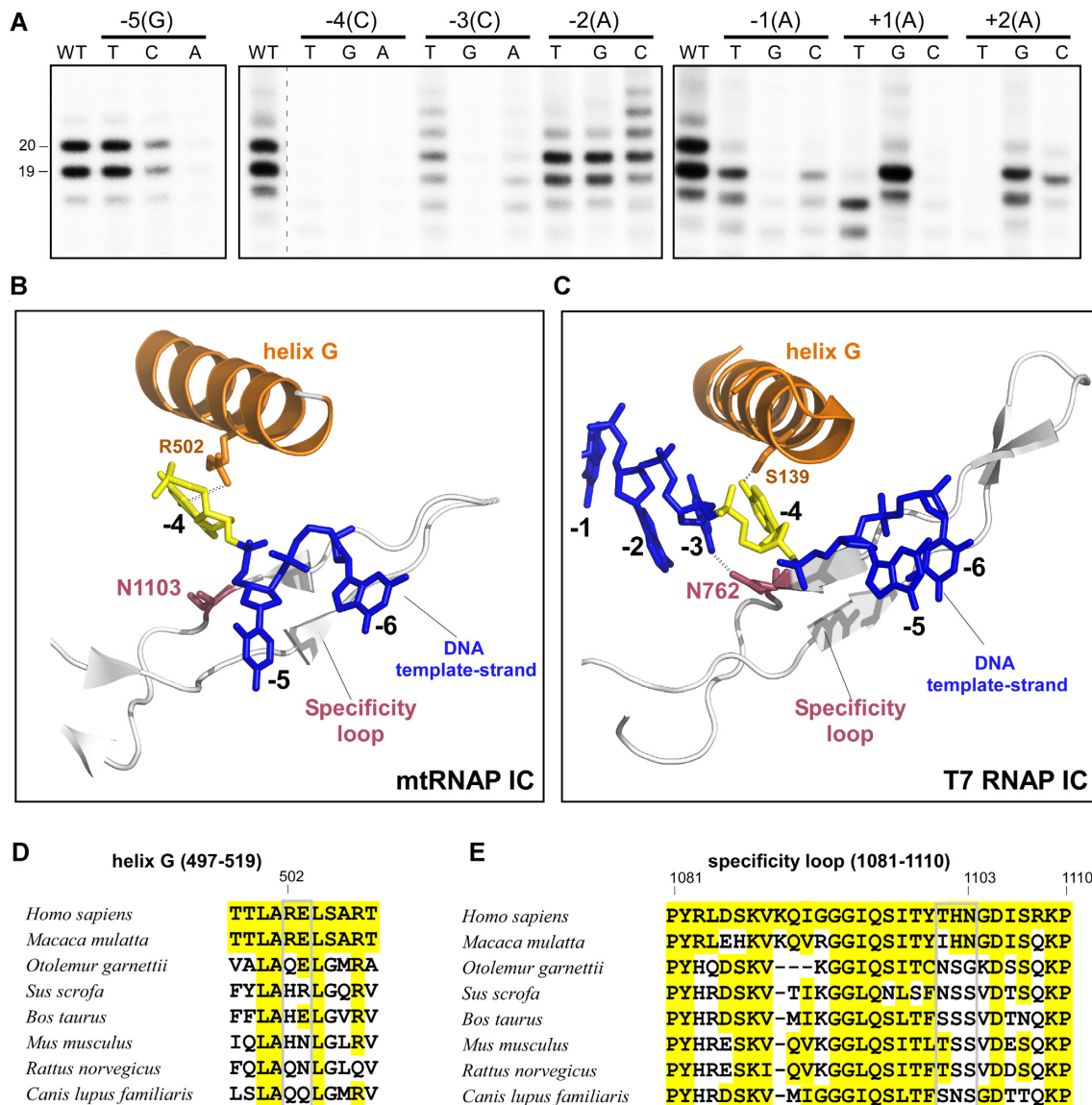
Analysis of the structure of the human mitochondrial IC reveals two structural elements that can specifically interact with the -3G and -4G bases in the template strand of DNA -





**Figure 2.** Structural features of mammalian mitochondrial promoters. (A) Mitochondrial promoters and their TSS in mammalian species. The phylogenetic tree of extant mammalian species is shown. Suborders, orders, and subfamilies of species, whose promoters have been identified, are highlighted with colored lines. The sequence logos represent the conservation between promoters in the species of the superorder Afrotheria, order Carnivora, Perissodactyla, Artiodactyla, Lagomorpha, and Chiroptera, suborder Hystricomorpha, infraorder Simiiformes, and the Murinae subfamily. Promoters have been identified for platypus (*Ornithorhynchus anatinus*, O.a.), koala (*Phascolarctos cinereus*, P.c.), and armadillo (*Dasypus novemcinctus*, D.n.), (logos are not shown). Species whose mitochondrial promoters were not identified are represented in grey lines. (B) Schematic structure of some promoter units in mammalian mtDNA. TSS of LSP and HSP and the distance between them are indicated, and the -3 and the -4 bases are highlighted. Repeated promoter units are indicated by the 4X and 2X signs in *Oryctolagus* (rabbits) and *Ochotona* (pikas) species. (C) Sequence alignment showing identical bases between LSP of *Equus caballus* (horse) and LSP of *Phyllostomus discolor* (pale spear-nosed bat). Yellow boxes highlight identical bases, TSS is indicated by arrows. (D) The structure of a palindromic LSP-HSP unit in dogs and wolves. The palindrome sequence is indicated in a blue box. (E) The structure of the LSP-HSP promoter in frogs (*Xenopus*). The palindrome sequence is indicated in blue boxes. (F) The structure of the LSP-HSP promoter in ducks (*Anas*). The palindrome sequence is indicated in blue boxes. (G) The interrupted palindromic promoter unit in some bat species. The palindrome sequence is indicated in blue boxes.





**Figure 3.** Promoter-mtRNAP interactions within transcription initiation complex. (A) Comprehensive mutagenesis of the LSP promoter. Run-off transcription assays were performed using synthetic LSP templates and human mtRNAP, TFB2M, and TFAM. (B, C) A close-up view of the active sites of human mtRNAP and T7 RNAP initiation complexes, PDB ID 6ERP and 1QLN. Conserved structural elements of the RNAPs – the G helix and the specificity loop-interact with the -3 and -4 bases of promoter DNA. Note that the residues in the proximity to -3 and -4 bases are shown as alanines in the SP loop of human IC. The position of the -5 base in the template strand of LSP is different from the -5 base in T7 RNAP IC because it has been artificially pre-melted. (D, E) Sequence alignment of the G helix (D) and the specificity loop (E) of mtRNAP in mammalian species. Residues that are identical or similar to human sequence are highlighted in yellow.

the G helix (residues 497–519) and the base of the SP loop (residues 1081–1110, Figure 3B). Of these two bases, only the density for the -4G base has been observed in the structure of the IC (3); however, the position of the -3G base can be modeled by homology modeling using the T7 RNAP IC (22). Comparison of the structures of the human mtRNAP and T7 RNAP initiation complexes reveals that the same structural elements in these polymerases interact with the analogous bases (Figure 3B and C). Residue S139 in the G helix of T7 RNAP hydrogen bonds to the -4 base, whereas the N762 residue in the SP loop recognizes the -3 base (Figure 3C). While sequence conservation of these structural el-

ements in mtRNAP and T7 RNAP is superficial, the N1103 residue in the SP loop and R502 in the G helix of mtRNAP are in a position to interact with the -3 and -4 bases in a manner that is analogous to that observed in T7 RNAP (Figure 3B).

Examination of sequence conservation in these regions of mtRNAP from various mammalian species reveals high variability in otherwise structurally conserved elements. The variable residues cluster around amino acid R502/E503 (indicated as ‘RE’) in the G helix (human mtRNAP numbering here and throughout the manuscript), and the substitutions include polar or charged residues such as HR, HE,

Table 2. Comprehensive mutagenesis of human LSP

WT base <sup>1,2</sup>	Base substitutions	Transcription efficiency <sup>3</sup> , % of WT	Notes
+4A (403)	T	19+/-12	
	G	38+/-8	
	C	22+/-12	
+3G (404)	T	28+/-4	
	C	13+/-2	
+2A (405)	A	<2	
	T	<2	
	G	42+/-15	
+1A (406)	C	24+/-10	
	T	<2	Compared to the 19 nt product in WT
	G	100+/-12	Compared to the 19 nt product in WT
-1A (407)	C	<2	Compared to the 19 nt product in WT
	T	44+/-16	
	G	<2	
-2A (408)	C	20+/-5	
	T	95+/-22	
	G	75+/-30	
-3C (409)	C	20+/-6	Enhanced transcript slippage
	T	5+/-2	Enhanced transcript slippage
	G	<2	
-4C (410)	A	<2	
	T	<2	
	G	<2	
-5G (411)	A	<2	
	T	100 +/-11	
	C	59+/-36	
-6C (412)	A	11+/-1	
	T	92+/-4	
	G	42+/-6	
-7C (413)	A	92+/-10	
	T	54+/-18	
	G	34+/-12	
-8A (414)	A	55+/-9	
	T	59+/-8	
	G	102+/-17	
-9T (415)	C	80+/-33	
	G	48+/-5	
	C	71+/-3	
-10A (416)	A	51+/-5	
	T	105+/-3	
	G	97+/-7	
	C	98+/-3	

<sup>1</sup>In the non-template strand of the LSP promoter.

<sup>2</sup>Position in human mtDNA is given in parenthesis.

<sup>3</sup>Assayed by the run-off transcription assay using all four NTPs and measured for the 19 and 20 nt long RNA products.

HN, QQ, QN (Figure 3D). Similarly, in the cluster of the polar/charged residues in the base of the SP loop of human mtRNAP - T1101/H1102/N1103 (or 'THN'), variations involving other residues capable of forming hydrogen bonds with DNA are found (Figure 3E). They include the following variants: NSS, NHS, SSS, TSS, SNS, SQS, and some others. The combination of these variable residues in both the G helix and the SP loop can, in principle, serve as a basis for recognition of the -3 and -4 bases in promoter DNA.

### *In vitro* reconstitution of the porcine mitochondrial transcription system

To understand further the molecular basis of mitochondrial promoter recognition, we cloned and purified porcine mtRNAP, TFB2M and TFAM and reconstituted the porcine *in vitro* transcription system using the control region of *Sus scrofa* mtDNA (Supplementary Figure S7A). Human and porcine promoters share no sequence homology (Figure 4A), allowing probing species-specific interactions using transcription assays. Unlike human mitochondrial promoters, porcine LSP and HSP contain identical 10 bp regions spanning the -5 to +5 interval (Figure 4A). Similar to human promoters, porcine LSP and HSP possess limited homology in the putative TFAM binding region (Figure 4A).

We obtained robust transcription using two fragments of porcine mtDNA that encompass LSP (position 1011–1139) and HSP (1064–1175). As is the case for human transcription (4), porcine transcription critically depends on the simultaneous presence of transcription initiation factors TFAM and TFB2M, which act synergistically (Figure 4B,C). Using the appropriate RNA markers and [ $\alpha$ -<sup>32</sup>P] ATP, we identified the exact location of the TSS in both LSP and HSP (Figure 4A-C), in full agreement with our SRA database analysis (Table 1).

Human and porcine transcription machinery display remarkable specificity towards the autologous promoters, as no specific transcripts were detected when heterologous promoters were used (Figure 4D,E). This high specificity of recognition is unlikely defined by interactions of TFAM with the promoter DNA, as substitution of human TFAM with its porcine homolog, or *vice versa*, did not result in a significant reduction in transcription (Figure 4F, Supplementary Figure S7B–S7D). Analysis of TFAM-mtRNAP interacting interface, which involves the C-terminal 'tail' of TFAM and the 'tether' helix of mtRNAP (3) (Figure 4G), reveals its conservation in many mammalian species. Residues in TFAM involved in hydrogen-bonding with mtRNAP - R159, R210, and E214 - are identical in human, mouse, and porcine TFAM, explaining their activity in the transcription reactions above (Figure 4H).

In contrast to TFAM, the substitution of human TFB2M with its porcine counterpart resulted in an almost complete loss of transcription activity on both promoters (Figure 4I, Supplementary Figure S7E). Similarly, human TFB2M was a poor substitute for porcine TFB2M, though some residual activity was detected at a high concentration of this protein (Figure 4J). TFB2M-mtRNAP interaction appears to be broadly conserved in mammals, with the notable exception of a salt bridge between the B-loop of mtRNAP and the C-terminal domain in TFB2M. Disruption of R601-D346 interaction in human mtRNAP results in a 5-fold reduction of transcription (3). In both human and porcine TFB2M, however, the D346 residue is preserved, suggesting that the dramatic decrease in transcription activity observed in experiments in Figure 4I could be due, at least in part, to incompatibility of the human promoter with porcine TFB2M. Finally, the substitution of human mtRNAP with porcine mtRNAP, or *vice versa*, resulted in a complete loss of transcription activity (Figure 4K and Supplementary Figure S7D).





### Species-specific recognition of the -3 base in promoter DNA

Structural data and the comprehensive mutagenesis analysis described above (Figure 3) suggest that the -3 and -4 bases in the human promoters are critical for their recognition. Notably, these bases differ between human and the newly identified porcine promoters (Figure 4). Similar to the situation with human promoters, substitutions of bases -3 and -4 in porcine LSP results in a significant decrease in transcription efficiency (Figure 5A), suggesting that recognition of these bases by mtRNAP could serve as a molecular basis for mitochondrial promoter specificity in different mammals. In agreement with this observation, the residues in the promoter-recognizing elements of mtRNAP located in the proximity to -3 and -4 bases (Figure 3B) are markedly divergent in mammalian species (Figure 3D,E), suggesting that they may play an essential role in species-specific promoter recognition (Figure 5B).

To investigate the mechanism of promoter recognition by mtRNAP, we substituted residues in the SP loop and the G helix of human mtRNAP that are in a position to interact with -3 and -4 DNA bases with their counterparts from porcine mtRNAP (Figure 5C) and assayed the effect of the substitutions on recognition of native and modified promoters. Three human mtRNAP variants were generated. The first variant (termed 'HR') involved substitutions of R502 and E503 to histidine and arginine, correspondingly, the residues found in the G helix of *Sus scrofa* mtRNAP (Figure 5B,C). Another mutant, 'NSS', involved the non-conserved residues in the specificity loop of mtRNAP - T1101N, H1102S, and N1103S. Finally, the third constructed variant, called 'pentamutant', combined the aforementioned mutations in helix G and the SP loop of mtRNAP (Figure 5C).

While these mutants retained full catalytic activity in a primer extension assay (Supplementary Figure S8A), they showed significantly reduced transcription during initiation on the human native promoter (Figure 5D, lanes 1–4). This confirms the critical role of these residues in promoter recognition.

When the mutant '-3T' promoter was used, the activity of WT mtRNAP was reduced ~5 fold (Figure 5D, lane 5). Remarkably, the mutant mtRNAPs demonstrated a higher transcription activity on the '-3T' promoter than the native promoter (Figure 5D, lanes 5–8). Thus, while the activity of the HR mutant increased only marginally, the NSS and the pentamutant mtRNAP showed a 2.5–4 fold increase, suggesting a clear preference in recognition of the 'porcine' -3 T-A base pair. Unexpectedly, none of these mutants were active on the '-4A' LSP promoter (Figure 5D, lanes 9–12) or the double mutant -3T/-4A LSP promoter (Supplementary Figure S8B), suggesting that additional mechanisms may be involved in recognition of this base. Recognition of the -3 base by the pentamutant was specific to the T-A base pair found in the porcine promoter, as no other base variation at this position produced efficient transcription (Figure 5E). Transcription of the '-3T' promoter was strictly TFB2M-dependent, confirming that the observed product was specific (Supplementary Figure S8C). These results indicate that the G-helix and the SP loop of mtRNAP harbor residues that define the specificity of recognition of the -3 base in mitochondrial promoters.

### TFB2M and mtRNAP recognize the -4 base of the promoter

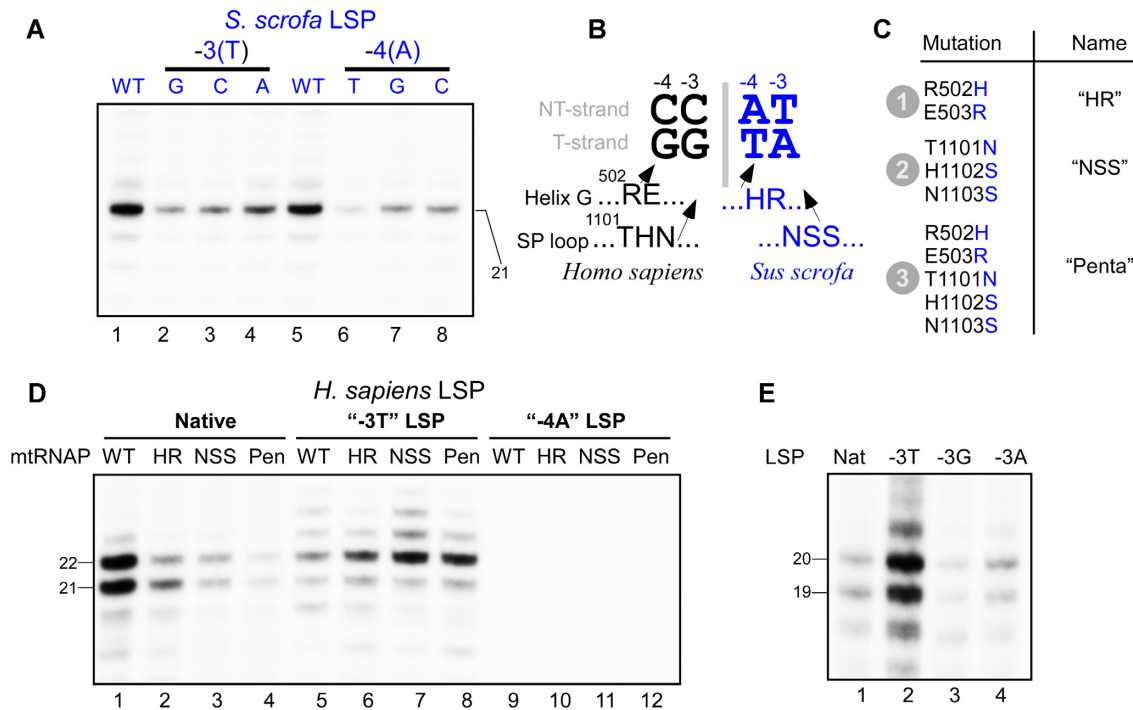
The finding that mutations in mtRNAP did not improve recognition of the 'porcine' -4A promoter variant suggested that the mechanism of recognition of this base pair involves additional interactions, possibly with TFB2M. Indeed, TFB2M makes extensive contacts with the non-template strand of promoter DNA during transcription initiation (3).

The -4 to +3 promoter region is melted in the IC, and TFB2M appears to bind the single-stranded DNA (3). We, therefore, can assess the effect of substitutions in a single-stranded context. To this end, to clarify whether recognition of the bases near the TSS occurs in the template (TS) or non-template strand (NT), we used a set of mismatched human LSP templates, in which a single base was substituted by the base found in the porcine promoter (Figure 6A). The substitution at the -3 TS position decreased transcription efficiency, while substitution of the -3 NT base had no effect (compare lanes 5 and 6, Figure 6A). This confirms our structural observations that the -3 base-pair is recognized in the template strand by mtRNAP (Figure 3B). The substitution at the -1 NT position also caused a decrease in transcription, suggesting that this base is being recognized (compare lanes 1 and 2, Figure 6A). Consistent with the data above (Figure 3A), mismatches at -2 did not result in significant changes in transcription efficiency, while the bases at position -4 appear to be recognized in both strands (Figure 6A; compare lanes 3,4 and 7,8). These data suggest that the -1 and -4 bases in the non-template strand are important for promoter recognition and are likely recognized by TFB2M, as no contacts of mtRNAP with the non-template strand in this region were observed in the IC structure (3).

To further demonstrate the ability of TFB2M to recognize bases -1 and -4, we compared the transcription efficiency of a modified porcine LSP in the presence of porcine or human TFB2M (Figure 6B). As expected, the substitution of base-pairs -1, -3, or -4 in the porcine LSP with those from the human LSP sequence resulted in decreased transcription in reactions involving porcine proteins. (Figure 6B, lanes 1–4). However, in the presence of human TFB2M, a notably increased transcription was observed in the case of substitution in the -1 base-pair, suggesting that it is specifically recognized (compare lanes 5 and 6, as opposed to 1 and 2, Figure 6B). A slight increase in transcription efficiency was observed in the case of the -4 base-pair (lane 5 vs. 8, as opposed to 1 vs. 4, Figure 6B). Again, no changes were observed for the -3 base (lane 3 and lane 7 vs its own controls), which is recognized by mtRNAP only.

To probe the selectivity of TFB2M interactions with the -1 and -4 non-template bases, we compared transcription efficiency using human TFB2M and the porcine LSP having all possible base-pair substitutions at position -1 and -4 (Figure 6C and Supplementary Figure S8D). In the case of the -1 base substitution, human TFB2M demonstrated a clear preference towards its 'native' base (dAMP) in the non-template strand (Figure 6C, lane 4). A preferable recognition of pyrimidine bases was observed at the -4 NT position (dCMP is the native base in human LSP). (Supplementary Figure S8D, E).





**Figure 5.** Promoter recognition by human mtRNAP. (A) The -3 and -4 bases in the porcine promoter are important for promoter recognition. Transcription assays were performed using porcine proteins and native promoter (lanes 1 and 5) or promoter variants having substitutions at the -3 and -4 positions (Lanes 2–4, 6–8). (B) Schematic drawing illustrating the non-conserved residues in the specificity loop and the G helix in human (black) and porcine (blue) mtRNAP implicated in recognition of the bases -3 and -4 in the template strand of promoter DNA. (C) Human mtRNAP variants having substitutions in the G-helix and the specificity loop. The substitutions were made to match the corresponding residues found in porcine mtRNAP. (D) The G helix and the SP loop of mtRNAP recognize the -3 base in LSP. Transcription assay was performed using the native, '-3T' or '-4A' human LSP with the WT or mtRNAP mutants as indicated. (E) The pentamutant mtRNAP specifically recognizes the '-3T' LSP. Transcription assays were performed using the native or mutant LSP, in which the -3 base was changed to all possible variants.

In the experiments described above (Figure 5D), mutations in human mtRNAP allowed for recognizing the 'porcine' -3 but not the -4 base in human LSP, even though the structural data suggest the proximity of this base to the mutated amino acids. Since TFB2M recognizes the -4 NT base, we probed whether the -4 base in the template strand is indeed recognized by mtRNAP (Figure 6D). When the -1 and -4 base-pairs in human LSP were substituted with the 'porcine' -1 and -4 bases, low transcription efficiency was observed, even though porcine TFB2M, which recognizes bases -4 and -1 in this template, was used instead of human TFB2M (Figure 6D, lane 1). A robust transcription was detected only when the -4 base in the template strand was reverted to match the native human base (dG) (lane 2, Figure 6D). This increase in transcription was not caused by the mismatch, as the -4 mutation in either strand suppresses transcription in the presence of human TFB2M (Figure 6A). These data suggest that the -4 base-pair is recognized by mtRNAP in the template strand and TFB2M in the non-template strand (Figure 6E).

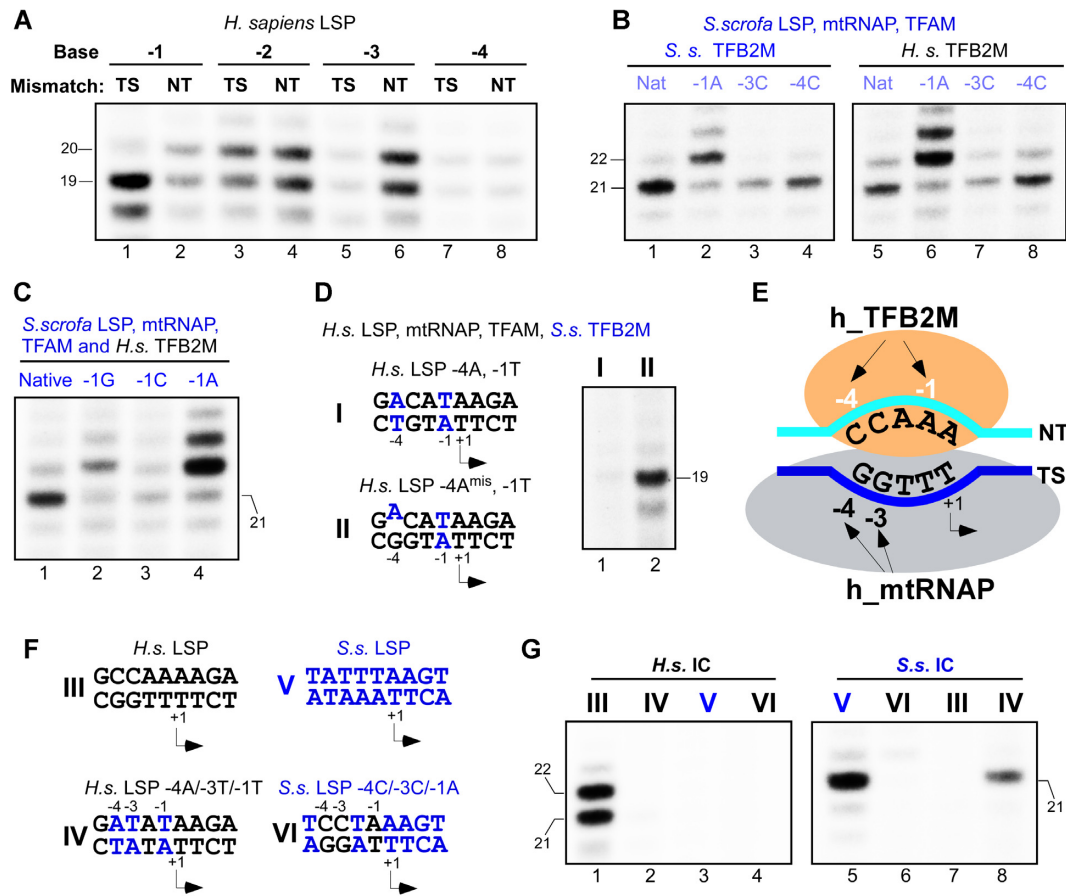
Unraveling the major principles of specific recognition of mitochondrial promoters (Figure 6E) challenged us to switch promoter specificity and force the porcine transcription machinery to recognize a modified human LSP and *vice versa* (Figure 6F,G). Substitution of the bases defining the specificity of the interactions (bases -1, -3, -4) in human LSP with their porcine counterparts resulted in complete inactivation of this promoter when human transcription machin-

ery was used (Figure 6G, left panel, lane 2). Similarly, 'human' mutations in porcine LSP inactivate transcription by porcine proteins (Figure 6G, right panel, lane 6). However, when the porcine transcription machinery was applied to the 'porcinized' human LSP, efficient and specific transcription activity was detected (Figure 6G, right panel, lane 8), indicating that just a few mutations generated in a mammalian mitochondrial promoter are sufficient to switch the specificity of recognition. The reciprocal approach did not result in a specificity switch (Figure 6G, left panel, lane 4). We speculate that an additional yet unknown sequence specificity determinant, which is not present in the human transcription machinery, may be required to recognize the porcine promoters.

## DISCUSSION

### Transcription start site in mammalian mitochondrial promoters

Definitive identification of the human TSS is one of the major findings of this work, as it paves the way to the analysis of the mechanisms behind species-specific promoter recognition in mtDNA. We have corrected the previously misidentified TSS for several model organisms such as humans, mice, rats, and pigs and, for the first time, revealed promoters for some avian species. Precise identification of the TSS for human and murine species is particularly important because



**Figure 6.** Mitochondrial promoters are recognized cooperatively by mtRNAP and TFB2M. (A) Mismatched nucleotides introduced into the NT DNA strand at positions -1 or -4 affect transcription efficiency. Transcription was performed using human proteins, and human LSP templates with single nucleotide mismatches at the position indicated. (B) Substitution of *S. scrofa* TFB2M with *H. s.* TFB2M compensates for the defects in the transcription of porcine LSP having mutations at positions -1 and -4. Transcription was performed using porcine mtRNAP and TFAM, and porcine (lanes 1–4) or human (lanes 5–8) TFB2M. LSP templates with single base-pair substitutions (porcine to human base) at the position indicated were used. Note the efficient transcript slippage observed on the -1A mutant template due to the generated stretch of three TMP bases in the template DNA strand (lanes 2 and 6). (C) Human TFB2M recognizes the dAMP base at the -1 position in the non-template strand. The native porcine LSP or its variants having all possible base-pair substitutions at position -1 were used in transcription assay with porcine TFAM, mtRNAP, and human TFB2M. (D) Human mtRNAP recognizes the -4 base in the template DNA strand. Human LSP template having base-pair substitutions at positions -1 and -4 (lane 1, template I) or a base pair substitution at position -1 and a mismatch at base -4 (lane 2, template II) were transcribed with human mtRNAP and TFAM, and porcine TFB2M. (E) Schematic illustration of promoter recognition by mtRNAP and TFB2M. MtRNAP recognizes the -3 and -4 bases in the template strand, while TFB2M recognizes the -1 and -4 bases in the non-template strand. (F, G) Porcine transcription machinery can recognize a modified human LSP. Transcription assays involved native human LSP (III), a porcinized human LSP with base-pair substitutions at positions -1, -3, and -4 (IV), native porcine LSP (V), and a modified (humanized) porcine LSP with base-pair substitutions at positions -1, -3 and -4 (VI). The promoter sequences near TSS are shown to the left (F). The left panel of the gel in (G) represents transcription reaction using human proteins, the right panel - transcription with porcine proteins.

the *in vitro* transcription systems have been previously developed. While the transcript slippage phenomenon complicated the TSS identification for the human promoters (23) and required the abovementioned experiments (Figure 1), the murine TSS also appears to be shifted 1–2 nt from the previously reported locations (24), as indicated by our RNAseq analysis. The high homology of the LSP promoters and the corresponding 5' ends of the transcripts in mice and rats brings high confidence into murine TSS identification by the RNAseq analysis (Supplemental Figure S4).

The observed 5' end heterogeneity of the human mitochondrial transcripts arises from the propensity of the mitochondrial machinery to engage in transcript slippage during transcription initiation (Figure 1). Transcript slippage has been demonstrated for various transcription systems

and plays an important role in gene expression and replication (19,25,26,27,28). Analysis of mitochondrial promoters reveals the conservation of the T-stretch near TSS, and as a consequence, the possibility of transcript slippage in LSP and HSP in humans, apes, and old-world monkeys but not in most other animals, which may have slippage in only one of the promoters or no slippage at all (Table 1, Supplementary Figures S2–S5). At this time, the significance of the conservation of transcript slippage mechanism in the mitochondria of humans and monkeys remains unclear.

Mitochondrial promoters in many mammalian species utilize ATP as a priming nucleotide, which is also the case for yeast mitochondrial promoters (29). The reason ATP is the preferred priming substrate is unclear. Pyrimidine bases do not appear to serve as efficient priming substrates, as

no transcription initiation was detected with UTP and CTP (Figure 3) and, consistently, no promoters having + 1 TMP or + 1 CMP have been found (Figure 2). However, our *in vitro* data demonstrate that initiation with GTP can be as efficient as with ATP. While promoters that initiate with GMP appear to be rare, we identified GMP at the 5' end of RNA transcripts from mice, rats, and a few other species (Figure 2, Supplementary Figures S2–S5).

### Promoter recognition in mammalian species

The core transcription system in mammals involves three proteins - mtRNAP, TFAM, and TFB2M - all of which have been implicated in recognizing promoter DNA. TFAM, a high mobility group protein, binds ~15 bp upstream of TSS and leaves a clear footprint on LSP and HSP in nuclease protection assay (30). However, the basis for such specific preference of promoter sequences is not clear, as most TFAM interactions occur in the minor groove of the DNA (31,32). Nevertheless, the initial binding of TFAM to the promoter, followed by recruitment of mtRNAP, plays an important role in transcription initiation by positioning the polymerase active site over the TSS during the formation of the pre-initiation complex (pre-IC). The alignment of promoter regions of mammalian species suggests a much higher degree of conservation of the TFAM HMG2 binding target than the HMG1 region (Figure 7A). This highlights the importance of the HMG2 region for the positioning of the C-terminus of TFAM, which recruits mtRNAP to the promoter. However, the detailed mechanism of HMG2 recognition by TFAM requires further investigation.

Because the DNA duplex is not melted in the pre-IC (8), it is possible that the initial recognition of the promoter involves major groove interactions with the SP loop of mtRNAP. The subsequent recruitment of TFB2M followed by promoter melting results in specific interactions with the unwound portion of the DNA. This includes sequence-specific interactions of mtRNAP with the -4 and -3 bases and TFB2M with the -1 and -4 bases (Figures 5 and 6).

Since the discovery of TFB2M (and its yeast analog Mtf1), it has been speculated that this protein plays a role similar to that of the sigma subunit in bacterial RNAP (12). Our data provide evidence that TFB2M is indeed involved in sequence-specific interactions within the initiation complex. These interactions likely occur after the initial promoter melting, as only the nucleotides in proximity to the TSS are recognized. Recognition of the non-template bases could be due to their flipping out and insertion into the pockets on the TFB2M surface, as observed in the yeast IC (33). Contrary to the SP loop's role in T7 RNAP (Cheetham et al., 1999), we detected only a moderate (10–50%) decrease in transcription efficiency when substitutions were introduced in the -6 to -12 promoter region. However, these regions of human promoters share no apparent sequence homology arguing against extensive sequence-specific contacts between the SP loop and major groove of DNA. Instead, the structure-based mutagenesis revealed interactions between the SP loop and residues at the C-terminal end of the G helix with the -3 base of the promoter. The amino acid composition in these structural elements is very

divergent in mammals and can serve as a recognition 'code' for promoter binding in different species. Indeed, analysis of the promoter regions in mammalian species suggests conservation of both -3 and -4 bases between LSP and HSP of the same species (Figure 7B). An example of a recognition code is observed in multiple *Chiroptera* species, where one of three possible combinations of the residues in mtRNAP strictly correlates with one of the -3 base variants in LSP (Figure 7C). An exception to this recognition rule is the species of the *Perissodactyla* order, such as horses, that lack conservation of the -3 or the -4 base between LSP and HSP (Figure 7B). We speculate that the mechanisms of promoter recognition in these species can be different and do not include recognition of the -3 and -4 bases. Instead, similar to yeast mtRNAP (34), the more upstream region of the promoter (from -7 to -5) could be recognized by the SP loop of mtRNAP in equine species (Supplementary Figure S5).

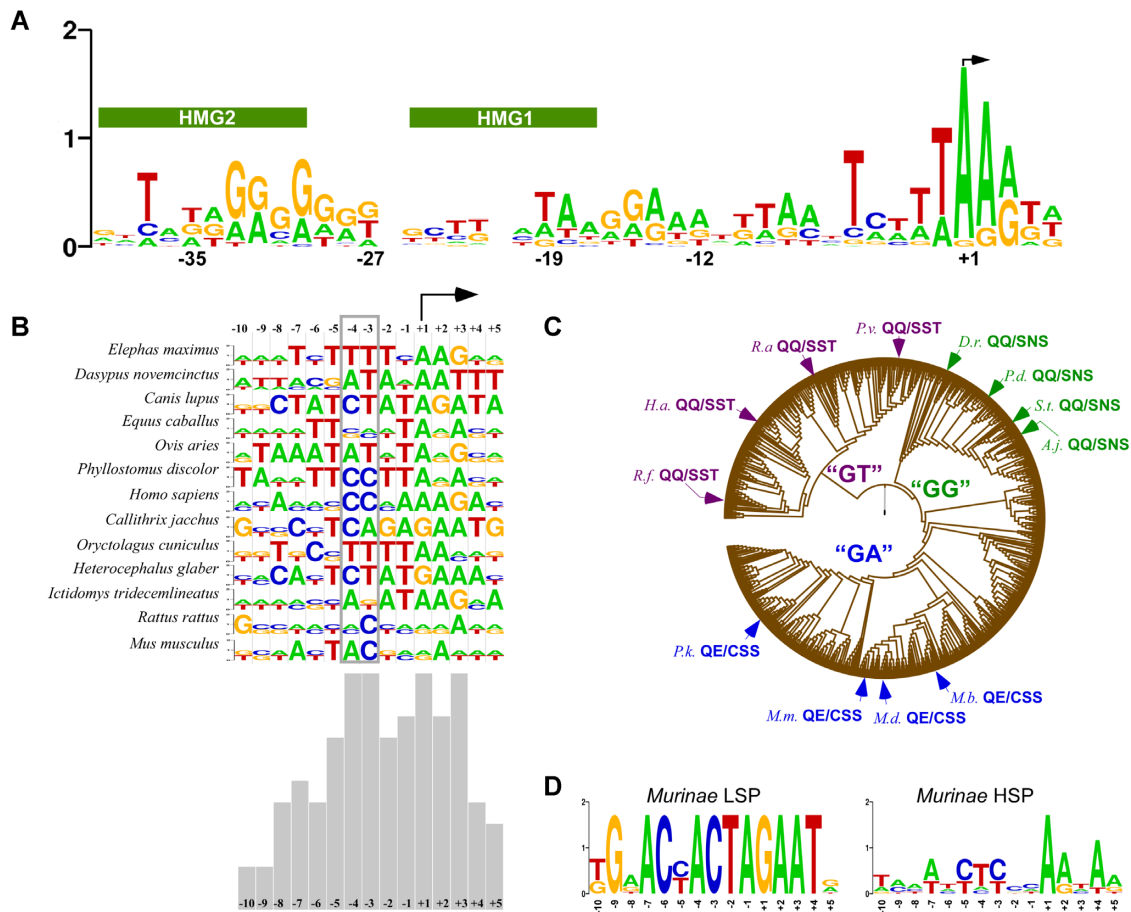
### Transcription machinery, mitonuclear incompatibility, and speciation

The interplay between the nuclear and the mitochondrial genome is crucial for cell homeostasis. Because of the much higher mutation rates for mtDNA, the nuclear genome must adapt rapidly to preserve mitochondrial functions. Thus, nuclear genes that have functional interactions with mitochondrial genes evolve faster than other nuclear genes (35). This mitonuclear coevolution is known to be a key factor in hybrid incompatibility, and as a consequence, an important step in speciation (36).

Since both nuclear-encoded mtRNAP and TFB2M recognize mitochondrial promoters, these genes would be expected to co-evolve with the mitochondrial DNA. Indeed, elevated evolutionary rates have been detected for both proteins. TFB2M is subject to positive selection in hominoids, Komodo dragons, and mole rats (37–39). Similarly, POLRMT shows a high rate of evolution and positive selection in the copepod *Tigriopus californicus* (40). The mtDNA of geographically isolated *Tigriopus species* is highly divergent (40), explaining the hybrid incompatibility observed in these species. The offspring produced from crosses between isolated populations of *T. californicus* preferentially inherit maternal mtRNAP along with their mtDNA (41). This suggests that other offspring variants, which inherit paternal mtRNAP and maternal mtDNA, are incompetent, likely because defects in promoter recognition are too severe and affect mitochondrial biogenesis.

The involvement of mtRNAP in hybrid incompatibility, which can lead to speciation, is further supported by an interesting finding in reptiles. As a result of an ancient geographic barrier, the mtDNA of two chameleon populations in Southern and Northern Israel significantly diverged (42). Now that the barrier is lost, both populations can mate and share nuclear genes, but the mitochondrial genome remains specific to each population. One of the nine polymorphisms identified in chameleon nuclear genes was mapped to the specificity loop (Q1090L) of mtRNAP. Interestingly, while two mtRNAP variants (Q and L) were present in both populations, they rarely appeared as homozygous in combination with mtDNA from the other population, indicat-





**Figure 7.** Evolution of the mammalian mitochondrial promoters. (A) Sequence conservation in TFAM-binding region of LSP and HSP in mammalian species. The sequence logos were built by alignment of promoters of all species for which TSS has been identified (Table 1), except the *Rodentia* species. (B) Sequence conservation between LSP and HSP in individual mammalian species as identified by RNAseq analysis. Top. The sequence logos were built by alignment of HSP and LSP in mtDNA of the same species. The -3 and -4 bases are indicated by a red box. Bottom. The schematic illustrates the conservation of the promoter bases in the species shown above. (C) Amino acid variation in the G helix and the SP loop of some bat species and the bases they recognize in LSP. The phylogenetic tree of the extant bat species is shown. Amino acids in the SP loop and the G helix of mtRNAP are indicated next to the species names. The recognized residues in the LSP promoters are indicated (bold caps) for the branches of the *Chiroptera* order. *Rhinolopus ferrumequinum* (*R.f.*), *Hipposideros armiger* (*H.a.*), *Rousettus aegyptiacus* (*R.a.*), *Pteropus vampyrus* (*P.v.*), *Desmodus rotundus* (*D.r.*), *Phyllostomus discolor* (*P.d.*), *Sturnira tildae* (*S.t.*), *Artibeus jamaicensis* (*A.j.*), *Myotis brandtii* (*M.b.*), *Myotis davidii* (*M.d.*), *Myotis myotis* (*M.m.*), *Pipistrellus kuhli* (*P.k.*). (D) Sequence conservation of LSP (left) and HSP (right) in species of the *Murinae* subfamily. The sequence logos for the -10/+5 promoter regions are shown.

ing that inheriting the ‘mismatched’ mtRNAP/mtDNA reduced the fitness of the individual (43). These findings suggest that mitochondrial promoters have diverged in these chameleon populations and that the observed polymorphism in the specificity loop of mtRNAP is an adaptive modification to this change. Unfortunately, the mitochondrial transcription system in reptiles has not been defined, precluding the analysis of promoter changes in these species.

Our findings suggest that the most dramatic effect on transcription initiation is related to the role of the -4 and -3 base in promoter recognition. The conservation of these bases is apparent between LSP and HSP of individual species, as both promoters are recognized by the same set of proteins (Figure 7B). However, transcription at LSP generates the mRNA and a replication primer (44). Therefore, mutations at these bases would produce transcription- and replication-incompetent genomes. At the same time, mutations of the -3 and -4 bases in the HSP promoter, which is

not involved in the replication primer synthesis, can give rise to replication-competent but transcription-incompetent, or ‘selfish’ mtDNA. Indeed, the HSP promoters are significantly more divergent in mammalian species with non-palindromic promoters, as seen with the *Murinae*, *Artiodactyla*, *Lagomorpha*, and *Perissodactyla* species. For example, in *Murinae*, LSP contains eleven absolutely conserved residues, while HSP – only one (Figure 7D). It is tempting to speculate that mutations in HSP that generate selfish mtDNA can drive the evolution of mitochondrial promoters. Further, identification of the TSS of more than 50% of mammals will enable studies of the evolution of the protein-DNA recognition mechanisms in mitochondria and a better understanding of mitonuclear coevolution.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.



## ACKNOWLEDGEMENTS

We thank the past and present members of the Temiakov laboratory. We thank Dr. W.T. McAllister for the critical reading of the manuscript and helpful discussion.

*Author Contributions:* A.Z.O. cloned and purified proteins, constructed mutants, performed transcription assays, and identified promoters, Y.M. isolated mitochondria, and performed primer-extension and transcription assays, A.S. purified proteins and performed transcription assays. A.S., A.Z.O., M.A., and D.T. analyzed the data, and D.T. wrote the manuscript.

## FUNDING

NIH R35 GM131832 (D.T.)

*Conflict of interest statement.* None declared.

## REFERENCES

- Tobler, M., Barts, N. and Greenway, R. (2019) Mitochondria and the origin of species: bridging genetic and ecological perspectives on speciation processes. *Integr. Comp. Biol.*, **59**, 900–911.
- Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P. and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase. *Nature*, **478**, 269–273.
- Hillen, H.S., Morozov, Y.I., Sarfallah, A., Temiakov, D. and Cramer, P. (2017) Structural basis of mitochondrial transcription initiation. *Cell*, **171**, 1072–1081.
- Litonin, D., Sologub, M., Shi, Y., Savkina, M., Anikin, M., Falkenberg, M., Gustafsson, C.M. and Temiakov, D. (2010) Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro. *J. Biol. Chem.*, **285**, 18129–18133.
- Morozov, Y.I. and Temiakov, D. (2016) Human mitochondrial transcription initiation complexes have similar topology on the light and heavy strand promoters. *J. Biol. Chem.*, **291**, 13432–13435.
- Morozov, Y.I., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P. and Temiakov, D. (2014) A novel intermediate in transcription initiation by human mitochondrial RNA polymerase. *Nucleic Acids Res.*, **42**, 3884–3893.
- Gaspari, M., Falkenberg, M., Larsson, N.G. and Gustafsson, C.M. (2004) The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J.*, **23**, 4606–4614.
- Shi, Y., Dierckx, A., Wanrooij, P.H., Wanrooij, S., Larsson, N.G., Wilhelmsson, L.M., Falkenberg, M. and Gustafsson, C.M. (2012) Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proc. Natl. Acad. Sci. USA*, **109**, 16510–16515.
- Posse, V., Hoberg, E., Dierckx, A., Shahzad, S., Koolmeister, C., Larsson, N.G., Wilhelmsson, L.M., Hallberg, B.M. and Gustafsson, C.M. (2014) The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation. *Nucleic Acids Res.*, **42**, 3638–3647.
- Morozov, Y.I., Parshin, A.V., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P. and Temiakov, D. (2015) A model for transcription initiation in human mitochondria. *Nucleic Acids Res.*, **43**, 3726–3735.
- Nayak, D., Guo, Q. and Sousa, R. (2009) A promoter recognition mechanism common to yeast mitochondrial and phage T7 RNA polymerases. *J. Biol. Chem.*, **284**, 13641–13647.
- Jang, S.H. and Jaehning, J.A. (1991) The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *J. Biol. Chem.*, **266**, 22671–22677.
- Chang, D.D. and Clayton, D.A. (1986) Precise assignment of the light-strand promoter of mouse mitochondrial DNA: a functional promoter consists of multiple upstream domains. *Mol. Cell. Biol.*, **6**, 3253–3261.
- Sologub, M., Litonin, D., Anikin, M., Mustaev, A. and Temiakov, D. (2009) TFB2 is a transient component of the catalytic site of the human mitochondrial RNA polymerase. *Cell*, **139**, 934–944.
- Sarfallah, A. and Temiakov, D. (2021) In vitro reconstitution of human mitochondrial transcription. *Methods Mol. Biol.*, **2192**, 35–41.
- Schwinghammer, K., Cheung, A.C., Morozov, Y.I., Agaronyan, K., Temiakov, D. and Cramer, P. (2013) Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.*, **20**, 1298–1303.
- Letunic, I. and Bork, P. (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.*, **47**, W256–W259.
- Upham, N.S., Esselstyn, J.A. and Jetz, W. (2019) Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.*, **17**, e3000494.
- Anikin, M., Molodtsov, M.V., Temiakov, D. and McAllister, W.T. (2010) Transcript slippage and recoding. In: Atkins, J. and Gesteland, R. (eds). *Recoding: Expansion of Decoding Rules Enriches Gene Expression, Nucleic Acids and Molecular Biology*. Vol. **24**. Springer, NY, pp. 409–432.
- Yang, S., Zheng, W., Yang, C., Zu, R., Ran, S., Wu, H., Mu, M., Sun, S., Zhang, N., Thorne, R.F. et al. (2021) Integrated analysis of hub genes and micRNAs in human placental tissues from in vitro fertilization-embryo transfer. *Front. Endocrinol. (Lausanne)*, **12**, 774997.
- Posse, V. and Gustafsson, C.M. (2017) Human mitochondrial transcription factor B2 is required for promoter melting during initiation of transcription. *J. Biol. Chem.*, **292**, 2637–2645.
- Cheetham, G.M., Jeruzalmi, D. and Steitz, T.A. (1999) Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature*, **399**, 80–83.
- Chang, D.D. and Clayton, D.A. (1984) Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell*, **36**, 635–643.
- Chang, D.D. and Clayton, D.A. (1986) Identification of primary transcriptional start sites of mouse mitochondrial DNA: accurate in vitro initiation of both heavy- and light-strand transcripts. *Mol. Cell. Biol.*, **6**, 1446–1453.
- Guo, H.C. and Roberts, J.W. (1990) Heterogeneous initiation due to slippage at the bacteriophage 82 late gene promoter in vitro. *Biochemistry*, **29**, 10702–10709.
- Linton, M.F., Raabe, M., Pierotti, V. and Young, S.G. (1997) Reading-frame restoration by transcriptional slippage at long stretches of adenine residues in mammalian cells. *J. Biol. Chem.*, **272**, 14127–14132.
- Molodtsov, V., Anikin, M. and McAllister, W.T. (2014) The presence of an RNA:DNA hybrid that is prone to slippage promotes termination by T7 RNA polymerase. *J. Mol. Biol.*, **426**, 3095–3107.
- Sarfallah, A., Zamudio-Ochoa, A., Anikin, M. and Temiakov, D. (2021) Mechanism of transcription initiation and primer generation at the mitochondrial replication origin oriL. *EMBO J.*, **40**, e107988.
- Amiott, E.A. and Jaehning, J.A. (2006) Mitochondrial transcription is regulated via an ATP ‘sensing’ mechanism that couples RNA abundance to respiration. *Mol. Cell*, **22**, 329–338.
- Dairaghi, D.J., Shadel, G.S. and Clayton, D.A. (1995) Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator. *J. Mol. Biol.*, **249**, 11–28.
- Rubio-Cosials, A., Sidow, J.F., Jimenez-Menendez, N., Fernandez-Millan, P., Montoya, J., Jacobs, H.T., Coll, M., Bernado, P. and Sola, M. (2011) Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.*, **18**, 1281–1289.
- Ngo, H.B., Kaiser, J.T. and Chan, D.C. (2011) The mitochondrial transcription and packaging factor tfam imposes a U-turn on mitochondrial DNA. *Nat. Struct. Mol. Biol.*, **18**, 1290–1296.
- De Wijnngaert, B., Sultana, S., Singh, A., Dharia, C., Vanbuel, H., Shen, J., Vasilchuk, D., Martinez, S.E., Kandiah, E., Patel, S.S. et al. (2021) Cryo-EM structures reveal transcription initiation steps by yeast mitochondrial RNA polymerase. *Mol. Cell*, **81**, 268–280.
- Biswas, T.K., Ticho, B. and Getz, G.S. (1987) In vitro characterization of the yeast mitochondrial promoter using single-base substitution mutants. *J. Biol. Chem.*, **262**, 13690–13696.
- Hill, G.E. (2020) Mitonuclear compensatory coevolution. *Trends Genet.*, **36**, 403–414.
- Ma, H., Marti Gutierrez, N., Morey, R., Van Dyken, C., Kang, E., Hayama, T., Lee, Y., Li, Y., Tippner-Hedges, R., Wolf, D.P. et al. (2016) Incompatibility between nuclear and mitochondrial genomes

- contributes to an interspecies reproductive barrier. *Cell Metab.*, **24**, 283–294.
37. Lind, A.L., Lai, Y.Y.Y., Mostovoy, Y., Holloway, A.K., Iannucci, A., Mak, A.C.Y., Fondi, M., Orlandini, V., Eckalbar, W.L., Milan, M. *et al.* (2019) Genome of the komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nat. Ecol. Evol.*, **3**, 1241–1252.
38. Sahn, A., Almada-Pagan, P., Bens, M., Mutalipassi, M., Lucas-Sanchez, A., de Costa Ruiz, J., Gorlach, M. and Cellerino, A. (2019) Analysis of the coding sequences of clownfish reveals molecular convergence in the evolution of lifespan. *BMC Evol. Biol.*, **19**, 89.
39. van der Lee, R., Wiel, L., van Dam, T.J.P. and Huynen, M.A. (2017) Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res.*, **45**, 10634–10648.
40. Barreto, F.S., Watson, E.T., Lima, T.G., Willett, C.S., Edmands, S., Li, W. and Burton, R.S. (2018) Genomic signatures of mitonuclear coevolution across populations of tigrionus californicus. *Nat. Ecol. Evol.*, **2**, 1250–1257.
41. Ellison, C.K. and Burton, R.S. (2008) Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proc. Natl. Acad. Sci. USA.*, **105**, 15831–15836.
42. Yaacov, Bar, Arbel-Thau, D., Zilka, K., Ovadia, Y., Bouskila, O. and Mishmar, D. (2012) Mitochondrial DNA variation, but not nuclear DNA, sharply divides morphologically identical chameleons along an ancient geographic barrier. *PLoS One*, **7**, e31372.
43. Bar-Yaacov, D., Hadjivasiliou, Z., Levin, L., Barshad, G., Zarivach, R., Bouskila, A. and Mishmar, D. (2015) Mitochondrial involvement in vertebrate speciation? The case of mito-nuclear genetic divergence in chameleons. *Genome Biol Evol.*, **7**, 3322–3336.
44. Clayton, D.A. (1982) Replication of animal mitochondrial DNA. *Cell*, **28**, 693–705.