

Complete sequence and gene organization of the mitochondrial genome of Siamensis Crocodile (*Crocodylus siamensis*)

Xuefeng Ji · Xiaobing Wu · Peng Yan ·
George Amato

Received: 26 October 2006 / Accepted: 17 January 2007 / Published online: 10 February 2007
© Springer Science+Business Media B.V. 2007

Abstract The complete sequence of mitochondrial genome of Siamensis Crocodile (*Crocodylus siamensis*) is determined by using PCR amplification, clone and primer-walking sequencing methods. The genome is 16836 bp in size, containing 13 protein-coding, 2 ribosomal and 22 transfer RNA genes. The mtDNA genome of Siamensis Crocodile is similar to most vertebrates in gene component, order, orientation, tRNA structure, low percentage of guanine and high percentage of thymine. The nonstandard stop codes (T) in two protein genes (Cox III and Cyt *b*) are more than those of most vertebrates. Transfer RNA genes range in length from 61 to 76 nucleotides, and their planar structure present characteristic clover leaf, except for tRNA-Ser (AGY) because of lacking the “DHU” arm.

Keywords *Crocodylus siamensis* · Genome · Mitochondrial · Structure

Introduction

The animal mitochondria gene content is highly conserved, which usually encodes 13 proteins, 2 ribosomal

RNAs (rRNAs), and 22 transfer RNAs (tRNAs). The gene order is also highly conserved among most vertebrates [1]. However, gene loss and gene rearrangement have been found in some taxa [2]. Comparisons of mitochondrial systems are useful for modeling genome evolution and phylogenetic inference [3]. They include gene content and gene arrangement, base composition, modes of replication and transcription, protein, tRNA and rRNA gene secondary structures, and genetic codon variations. These features are currently more accessible for study in the much smaller and simpler mitochondrial genomes [3, 4]. During the last 10 years, mitochondrial genome sequence and gene arrangement comparisons were employed as powerful new tools for resolving ancient phylogenetic relationships [2, 3, 5–7].

According to the traditional classification based on morphology, Crocodylia is divided into three major groups: the Family Alligatoridae (*Alligator*, *Caiman*, *Melanosuchus*, and *Paleosuchus*), the Family Crocodylidae (*Crocodylus*, *Osteolaemus*, *Tomistoma*), and the Family Gavialidae, which includes only one species, *Gavialis gangeticus* [8]. *Crocodylus siamensis* belongs to genus *Crocodylus*, Family Crocodylidae. As a freshwater crocodylian (a group that also includes alligators, caimans and the gharial), *Crocodylus siamensis* is one of the most endangered crocodiles in the wild, which is widely bred in captivity [9]. At present there are seven crocodylian species with their mitochondrial genomes having been completely sequenced [6, 7, 10–13]. But the data is not enough to explain the phylogenetic relationships of crocodylians. So we sequenced the complete sequence of mitochondrial genome of *Crocodylus siamensis* and analyzed its gene organization.

X. Ji · X. Wu (✉) · P. Yan
College of Life Sciences, Anhui Key Laboratory
of Conservation and Exploration for Important Biological
Resource, Anhui Normal University, Wuhu 241000, China
e-mail: wuxb@mail.ahnu.edu.cn

G. Amato
Molecular Systematics Laboratory, American Museum
of Natural History, New York, NY, USA

Materials and methods

Samples and sequencing

Crocodylus siamensis samples were obtained from the specimen storeroom of College of Life Sciences in Anhui Normal University. The fresh blood was stored at -80°C .

Mitochondrial DNA was extracted following the procedure described in Arnason et al. [14]. Extracted DNA was diluted 10 times and stored at -20°C until use as a template for polymerase chain reaction (PCR). The primers for PCR and sequencing were designed based on the complete mtDNA sequences of *Alligator mississippiensis* (The GenBank accession number is Y13113), *Alligator sinensis* (The GenBank accession number is AF511507) and *Caiman crocodilus* (The GenBank accession number is AJ404872). By using ClustalX 1.8 [15] we designed 66 primers (Table 1) for

the PCR amplification (synthesized by Shanghai Sangon Biotechnology Co., Ltd., Shanghai, China). The length of amplified products is from 900 bp to 1200 bp. All overlaps between the products were confirmed by direct sequencing of PCR products amplified from the purified mtDNA template. The whole mitochondrial genome was read at least two times.

PCR reactions were conducted on a PTC-200 thermal cycler with the following conditions: an initial denaturation step of 95°C (3 min) followed by 34 cycles of 95°C (30 s), $49\text{--}60^{\circ}\text{C}$ (30 s), and 72°C (90 s) followed by 72°C for 10 min. The PCR products were separated by electrophoresis in 1% agarose gels, then were purified by PCR Cleanup Kit (V-gene, Biotechnology Limited, Hangzhou, China) and sequenced with ABI 3730 (Shanghai Sangon Biotechnology Co., Ltd., Shanghai, China). At last we got the complete mtDNA sequences of *Crocodylus siamensis*. The determined nucleotide sequences are deposited

Table 1 Sequencing primers used in the analysis of *Crocodylus niloticus* and *Crocodylus siamensis* mitochondrial genomes

Primer name	Upper primer sequence (5' → 3')	Lower primer sequence (5' → 3')	Anneal temperature (°C)
Pn1	AATATGTACACACCGCCCGTC	ATTACGCTACCTTYGCACG	53.0
Pn3	GGCCTAACATAACGAATGACT	CGTTTGCCGAARGCAGAT	52.0
Pn6	TCTGGTTATCACTCTCACG	GACTAAATCRAGCTTTCGC	50.9
Pn8	CACTAGCAGAAACAAACCG	CTTCTGGTACTCAGAAGTG	52.0
Pn9	GCATAAACTTCCACCCCTC	GTTGAARKGATCCTCGTTC	50.0
Pn13	CGACCTCGATGTTGGATCAG	ATAGTGTGAACAGCCCTGC	54.0
Pn16	GCTCAACTTACACATGCA	TTCCAAGCACACCTTCCG	50.0
Pn19	CTTCCCACAACACTTCCT	TTTCTTGGACGTCTGAGG	53.0
nle1	CAGCGCAACCTAACACGAT	TCACGGACGTGAATGAGTGA	54.0
nle3	CCTAATCATCTTCTCATGGAC	GGTCTCTTGTTTTGTAAATGGC	53.0
nle4	AATCTGGGCGAGCTACTAGC	CCTTAGGTAGAGGTTGAGTC	54.0
nle5	GACCAGTCCGACTTTCATGT	GACTTAACGCCCGTGTCTGT	53.0
nle12	GATTGGGTCATTAATCTCCATA	TTCAMGGTCAGGGTCATGTTTT	52.0
nle14	CAACCTGCTTAACAACCGC	GGAGACTCATAGTAATGTG	51.0
nle15	ATGGTCTTTCAGCGATCCT	ATGCCTACTCCYTCTCAGCC	54.0
nle17	CACTGCTAATAACCAACGA	TGGCTGAGCTCTGTTCCGA	53.5
nle18	CATATCACCAGGGCTATCAAC	TATGGAGATTAATGACCCAATC	52.0
nle20	CTGTACCAATACAGCACTTCC	ATTATGGCAGCCATAGCCCC	56.0
nle21	CATTAGCTTCACGCTAGACA	GTGGTTGAGCCTTATTGGT	52.0
nle23	CTCAATCTATTGATGAGGCTC	TGTGACATGGAACCTTTGGCT	52.0
nle24	TATCCTGCTGACTACTACCA	GAGGTGATGAATGCGCTTA	53.0
cp4	CCCACCTGATAGACCAAACA	GCGTTCGTGGAGGAAGAT	52.0
cp5	CTTCCCACAACACTTCCT	AGTACTGTTGCTAGCCA	53.0
cp6	TTAGCCTACTCATCCGACC	TGGTTCGAGGTGAAGTG	51.0
L16	GAACCATTCGGCTATATAGGA	GGCTCTTCGTAGGTGTGAT	53.1
L17	CCTACGGATCAACATTC	GTCAATAGGGAGCCATGA	54.0
L18	ACCATGCCAACATCAGCCAAGCT	GTCTAAAGTCCAAGTGAG	50.0
P78	CGCCTGTTTACCAAAAACAT	ACGTGATACTGAGTTCAGA	53.0
CS1	GAGAATTACGAGCAACTTGC	CATCTTCCCTTGC GG TACT	53.0
CS3	GAGCCTATTCTTTTACCTACGA	GTTGGTATAGGATGGGGTCTCC	54.0
CS4	GATACCACGACGATATTCC	GATCACTTATGTCCTGGC	52.0
CS5	TACGAACGAAACACTCACG	GAGTGTAAATAGGGCTGATAC	52.0
CS6	CCTAAAACCTCTGCTCATGC	CTGGTTAATGTGGTCTGGT	52.0

Respectively and Y = C or T, R = A or G, K = G or T, M = A or C

in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession number DQ353946.

Gene identification and sequence analysis

With the analysis of DNASTar (Version 5.01) and Sequin (Version 5.35), 12 Heavy-strand encoded protein-coding genes and one Light-strand encoded protein-coding gene were obtained. By the software of tRNA Scan-SE1.21 (http://lowelab.ucsc.edu/tRNA_Scan-SE), 21 transfer RNA (tRNA) genes were found. The sequences were aligned by using ClustalX 1.8 [15], and then were corrected by using DNASIS 3.5 (Hitachi). After checking the position of the tRNA-Ser (AGC) and rRNA genes, the positions of the rest genes were revised and determined.

Results and discussion

Characteristics of the crocodylian mitochondrial genomes

The total length of the mtDNA molecular is 16836 bp. 31 protein-coding genes, 22 tRNA genes, 2 rRNA genes, and one control region were identified in the mitochondrial genome of *Crocodylus siamensis* (Fig. 1).

The overall base composition of the H-strand is A: 32.01% (5389 bp); T: 24.76% (4168 bp); C: 28.47% (4793 bp); G: 14.77% (2486 bp). The relative order of nucleotide composition is A > C > T > G, with A being the most represented base (32.12%) and the least of G (14.74%). Similar to cases reported in other vertebrate mitochondrial genomes [16, 17], the base compositions of mtDNAs were skewed, with more A–T base pairs (56.77%) than G–C (43.23%) base pairs. The skew of AT is 0.13, which is higher than that of GC (−0.32). As other vertebrate mtDNAs, overlap and noncoding bases in some genes were also observed. In *Crocodylus siamensis*, the total size of the overlap of six times is 71 bp. Of the 71 bp, the ATPase 8 gene and ATPase 6 gene shared the 22 bp overlap, and NADH5 and NADH6 overlap by 44 bp. The total length of noncoding spacer nucleotides is 222 bp. The longest noncoding region of 56 bp, is located between Cyt *b* and tRNA-Thr.

Protein-coding genes

The 13 protein-coding genes in *Crocodylus siamensis* are 11380 bp. The longest one is NADH5 (1854 bp)

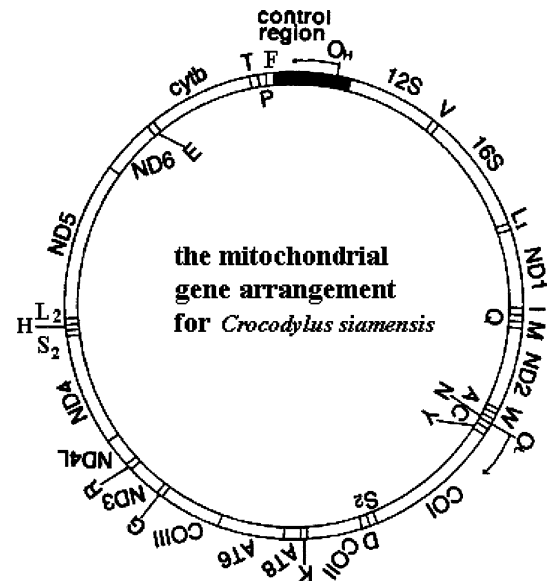


Fig. 1 Gene organizations of vertebrate mtDNAs. The organizations found in *Crocodylus siamensis* were shown in Table 2. Genes encoded by the heavy strand are shown outside the circle, whereas those encoded by the light strand are shown inside the circle. Gene abbreviations used are 12S, 12S rRNA; 16S, 16S rRNA; ND1–6, NADH dehydrogenase subunits 1–6; COI–III, cytochrome oxidase subunits I–III; AT6 and AT8, ATPase subunits 6 and 8; cyt *b*, cytochrome *b*; and one-letter codes of amino acids, tRNA genes specifying them (L1 and L2 for leucine tRNA genes specifying, respectively, UUR and CUN codons and S1 and S2 for serine tRNA genes specifying, respectively, AGY and UCN codons). O_H and O_L stand for the heavy-strand replication origin and the light-strand replication origin, respectively

gene, whereas the shortest one is ATPase 8 (162 bp) gene.

In *Crocodylus siamensis* mtDNA, ATG is the translation initiation codon in 7 of the 13 protein-coding genes. The start codon ATA occurs in the NADH1, NADH3, NADH5, while COX I and NADH4L begin with the nonstandard start codons GTG and ACC, respectively. In other crocodylians sequenced, the nonstandard start codons (ACC, AGC, ACT, ACA, ATC) occur in the NADH4L, which is a probable result of highly evolving rate. Standard termination codon TAA occur in 10 genes, however, AGG is found in NADH1. Incomplete termination codon T occurs in COX III and Cyt *b*.

tRNA and rRNA gene

Sequences of all 22 tRNA genes can be folded into a canonical cloverleaf secondary structure with the exception of tRNA-Ser (AGY), which loses “DHU” arm. These tRNAs range in size from 61 to 76 nucle-

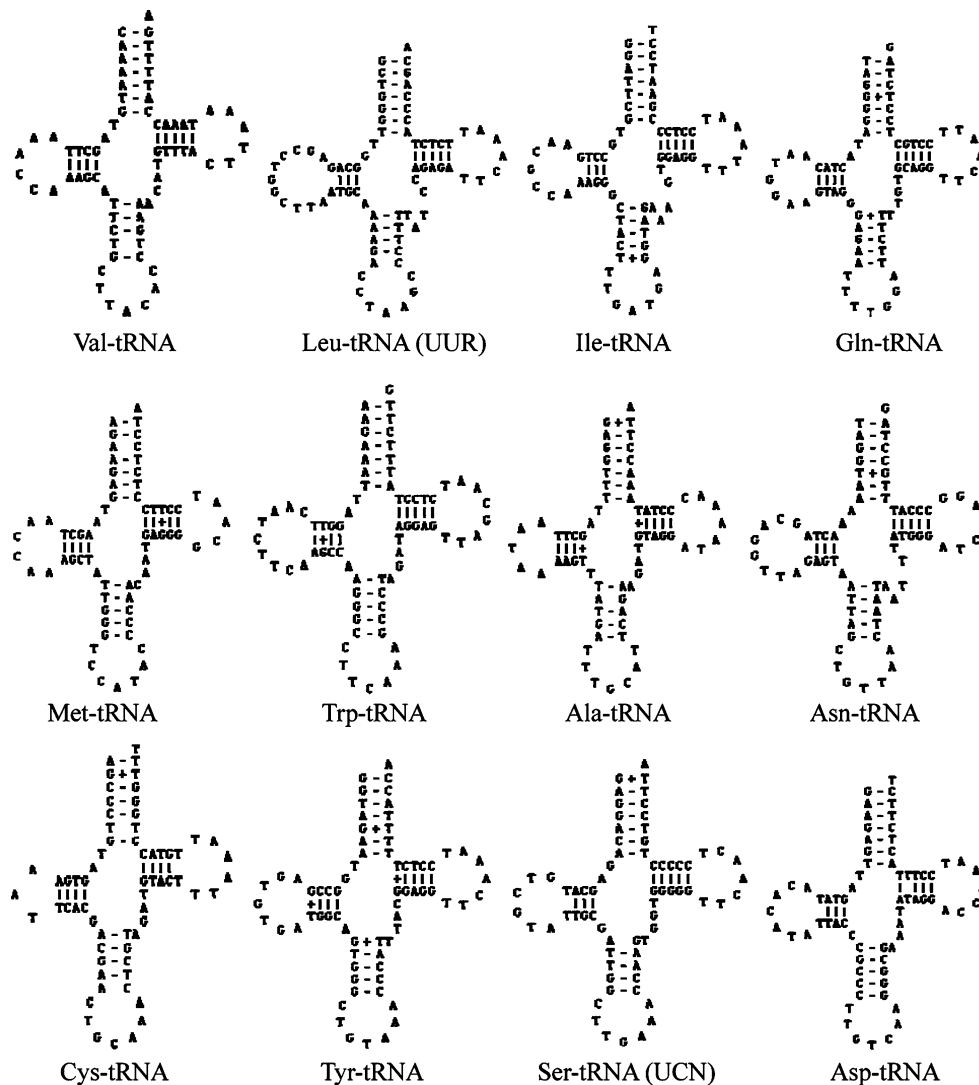


Fig. 2 Inferred secondary structure of tRNA in *Crocodylus siamensis*

otides. Figure 2 shows all the secondary structures of 22 tRNAs.

Noncoding sequences

The control region of mtDNA in *Crocodylus siamensis* is 1108 bp, which locates between the tRNA-Phe and 12S rRNA genes (Fig. 1). The general structure and conserved sequences of the crocodilian mitochondrial control region were comprehensively analyzed by Ray and Densmore [18]. In crocodilians, domain II is the most conserved part of D-loop and contains several conserved sequence boxes characterize in other vertebrates; domain III is the most variable part of D-loop and contains

several interesting sequence motifs, including tandemly repeated sequences and a long poly-A (*Crocodylidae*) or poly-C (*Alligatoridae*); domain I tending to be shorter than the same region in mammals and birds, contains sequences similar in structure to both the goose-hairpin and termination associated sequences (TAS), so it is more conservative than domain III.

The origin of L-strand replication (OLR), which usually locates in a cluster of five tRNA genes: tRNA^{Trp}-tRNA^{Ala}-tRNA^{Asn}-tRNA^{Cys}-tRNA^{Tyr} (WAN-CY) in vertebrates, lacks in the mtDNA of *Crocodylus siamensis*. Thus, the absence of OLR appears to be a common characteristic of all crocodilians. The OLR region between Asn-tRNA and Cys-tRNA, which has

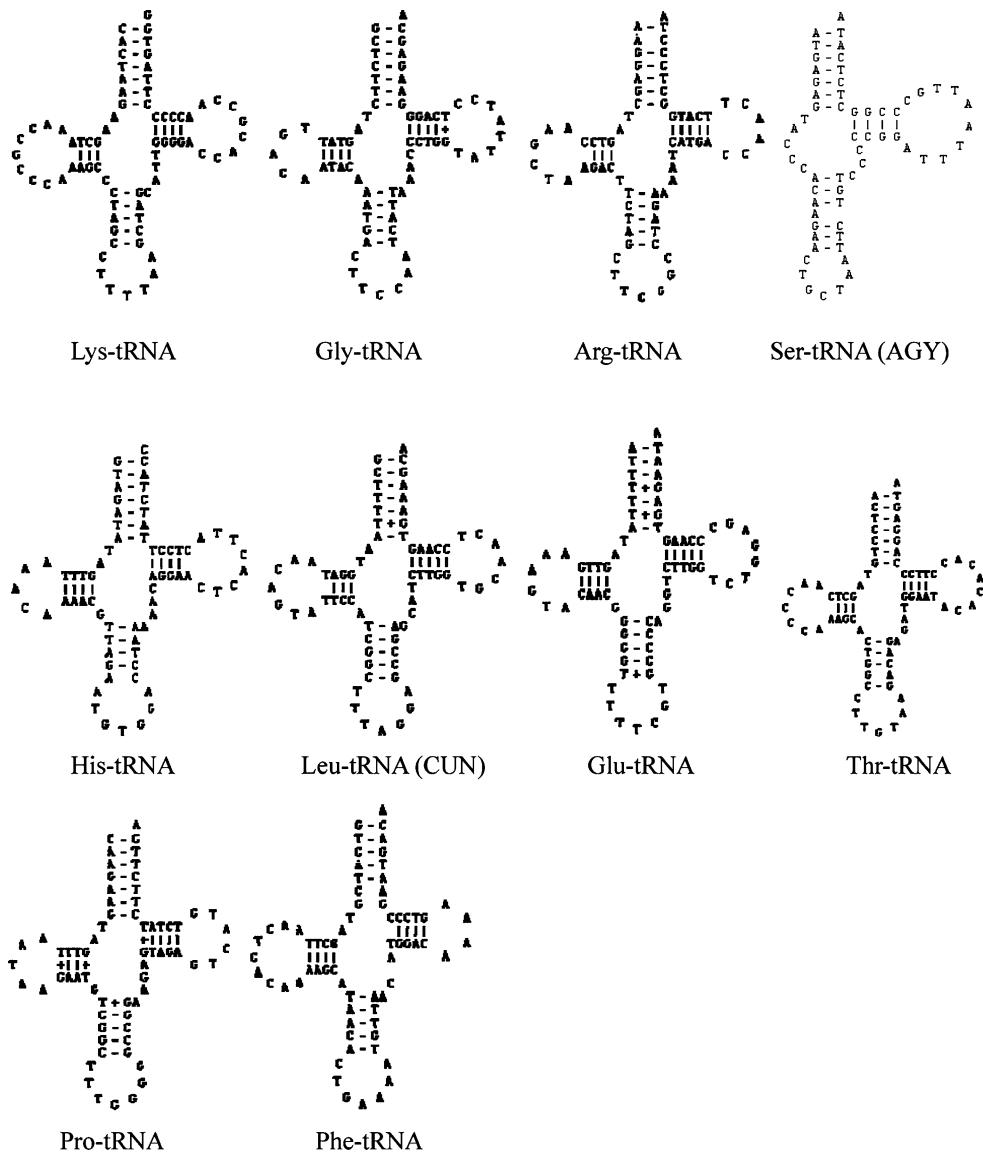


Fig. 2 continued

the potential to fold into a stable stem-loop secondary structure, only has “AATATT” 6 bp in *Crocodylus siamensis*. The absence of OLR is probable the result of rearrangement of tRNA.

The control region of *Crocodylus siamensis* mtDNA is similar with the others vertebrates, which is a most variable part of the mitochondrial genome. But this variable part still has some typical characteristic, for instance, microsatellite sequences and repetitive sequences (5'-CAACCTAGGCCAAAATAGGAA

GAAATTTTAAAAATTTT-3', 39 bp repeat four times). These features, which found in the control region are the origin of heavy strand (H-strand) replication and the promoters for both heavy and light strand (L-strand) transcription [19–21].

Acknowledgements This work was supported by National Natural Science Foundation of China (NSFC, No. 30470244), the Foundations for Excellent Youth in Anhui Province (04043409) and Key Laboratory of Biotic Environment and Ecological safety in Anhui Province.

Table 2 Locations of genes and noncoding regions in the mitochondrial genome of *Crocodylus siamensis*

Gene/region	Nucleotide no.		Size (bp)	aa
	From	to		
12S rRNA	1	985	985	
tRNA-Val	986	1054	69	
16S rRNA	1055	2651	1597	
tRNA-Leu(UUR)	2653	2728	76	
NADH1	2729	3691	963	320
tRNA-Ile	3690	3761	72	
tRNA-Gln (L)	3760	3830	71	
tRNA-Met	3833	3900	68	
NADH2	3905	4960	1056	351
tRNA-Trp	4961	5034	74	
tRNA-Ala(L)	5037	5105	69	
tRNA-Asn(L)	5107	5179	73	
tRNA-Cys(L)	5186	5251	66	
tRNA-Tyr(L)	5252	5322	71	
Cox I	5324	6880	1557	518
tRNA-Ser (UCN) (L)	6901	6970	70	
tRNA-Asp	6974	7044	71	
Cox II	7045	7728	684	227
tRNA-Lys	7730	7803	74	
ATPase 8	7805	7966	162	53
ATPase 6	7945	8640	696	231
Cox III	8640	9423	784	261
tRNA-Gly	9424	9493	70	
NADH3	9494	9841	348	115
tRNA-Arg	9852	9920	69	
NADH4L	9924	10217	294	97
NADH4	10211	11584	1374	457
tRNA-Ser(AGY)	11614	11674	61	
tRNA-His	11692	11761	70	
tRNA-Leu(CUN)	11789	11859	71	
NADH5	11863	13716	1854	617
NADH6(L)	13673	14200	528	175
tRNA-Glu(L)	14201	14271	71	
Cyt b	14287	15442	1156	385
tRNA-Thr	15499	15570	72	
tRNA-Pro(L)	15582	15649	68	
tRNA-Phe	15660	15728	69	
D-loop	15729	16836	1108	

References

- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27(8):1767–1780
- Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP (2003) Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* 29(2):289–297
- Boore JL (2004) Complete mitochondrial genome sequence of *Urechis caupo*, a representative of the phylum Echiura. *BMC Genomics* 5(1):67
- Garesse R, Carrodegua JA, Santiago J, Perez ML, Marco R, Vallejo CG (1997) *Artemia* mitochondrial genome: molecular biology and evolutive considerations. *Comp Biochem Physiol B Biochem Mol Biol* 117(3):357–366
- Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc Natl Acad Sci USA* 95(24):14226–14231
- Janke A, Erpenbeck D, Nilsson M, Arnason U (2001) The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. *Proc Biol Sci* 268(1467):623–631
- Wu XB, Wang YQ, Zhou KY, Zhu WQ, Nie JS, Wang CL (2003) The complete mitochondrial genome sequence of Chinese alligator and the phylogeny of crocodiles. *Chin Sci Bull* 48(18):1954–1958
- Mook CC (1921) Note on the postcranial skeleton in the Crocodilia. *Bull Am Mus Nat Hist* XLIV:67–100
- Temsiripong Y, P. Ratanakorn (2001) A review for reintroduction program of Siamese crocodile in Thailand. Proceedings of the Regional Working Meeting of the Crocodile Specialist Group, Guangzhou, China, IUCN-The World Conservation Union, Gland, Switzerland and Cambridge UK. Temsiripong, Y. Re-introduction of the Siamese crocodile. *CSG Newsletter*; 20(1), pp 10–13
- Janke A, Arnason U (1997) The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles). *Mol Biol Evol* 14:1266–1272
- Janke A, Gullberg A, Hughes S, Aggarwal RK, Arnason U (2005) Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodylians. *J Mol Evol* 61(5):620–626
- Ji XF, Wu XB, Li Y, Yan P, Amato G (2006) The mitochondrial genome of *Crocodylus niloticus* with implications for phylogenetic relationships among crocodylian species. *Acta Zool Sinica* 52(4):810–818
- Li Y, Wu XB, Ji XF, Yan P, Amato G (2006) The complete mitochondrial genome of the Salt-water Crocodile (*Crocodylus porosus*) and phylogeny of crocodiles. *Acta Genet Sinica* 2
- Arnason U, Gullberg A, Widegren B (1991) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J Mol Evol* 33(6):556–568
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25(24):4876–4882
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 141:173–216
- Asakawa S, Kumazawa Y, Araki T, Himeno H, Miura K, Watanabe K (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *J Mol Evol* 32:511–520
- Ray DA, Densmore L (2002) The Crocodylian mitochondrial control region: general structure, conserved sequences and evolutionary implications. *J Exp Zool* 294(4):334–345
- Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell* 28(4):693–705
- Clayton DA (1984) Transcription of the mammalian mitochondrial genome. *Annu Rev Biochem* 53:573–594
- Wong TW, Clayton DA (1985) In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. *Cell* 42(3):951–958