动 物 学 研 究 2005, Apr. 26 (2): 123 - 128 CN 53 - 1040/Q ISSN 0254 - 5853
Zoological Research

Complete Sequence and Gene Organization of the Mitochondrial Genome of Tokay (*Gekko gecko*)

HAN De-min[1, 2], ZHOU Kai-ya[1, *]

(1. *Jiangsu Key Laboratory for Bioresource Technology, College of Life Sciences, Nanjing Normal University, Nanjing* 210097, *China*; 2. *School of Life Science, Anhui University, Hefei* 230039, *China*)

Abstract: Long-PCR amplification, clone and primer-walking sequencing methods were employed in determine the complete sequence of mitochondrial genome of tokay (*Gekko gecko*). The genome is 16 435 bp in size, contains 13 protein-coding, 2 ribosomal and 22 transfer RNA genes. The mt genome of *Gekko* is similar to most of the vertebrates in gene components, order, orientation, tRNA structures, low percentage of guanine and high percentage of thymine, and skews of base GC and AT. Base A was preferred at third codon positions for protein genes is similar to amphibians and fishes rather than amnion vertebrates. The standard stop codes (TAA) present only in three protein genes, less than those of most vertebrates. Transfer RNA genes range in length from 63 to 76 nt, their planar structure present characteristic clover leaf, except for tRNA-Cys and tRNA-Ser (AGY) because of lacking the D arm.

Key words: *Gekko gecko*; Squamate; Complete sequence of mitochondrial genome; Gene organization

大壁虎线粒体基因组全序列及其结构

韩德民[1, 2], 周开亚[1, *]

(1. 南京师范大学 生命科学学院生物资源技术江苏省重点实验室, 江苏 南京 210097; 2. 安徽大学 生命科学学院, 安徽 合肥 230039)

摘要: 采用长 PCR 扩增、克隆和引物步行等方法, 测定了大壁虎(*Gekko gecko*)线粒体基因组全序列。序列全长 16 435 bp, 共有 13 个蛋白质编码基因、2 个 rRNA 基因和 22 个 tRNA 基因。基因组的组成、顺序、编码链的选择、tRNA 的结构、较低的碱基 G 含量、对碱基 T 的偏好以及 GC 和 AT 偏斜, 都与大部分脊椎动物相同或相近。但有些特征揭示了壁虎类的原始性: 蛋白质编码基因密码子第 3 位表现为对碱基 A 的偏好, 更接近两栖类和鱼类而不是羊膜动物; 标准终止密码子(TAA)只出现于 3 个蛋白质编码基因中, 比大部分脊椎动物少。tRNA 基因核苷酸长度为 63 ~ 76 nt, 除了 tRNACys 和 tRNASer(AGY)缺少 D 臂, 其余的二级结构均呈典型的三叶草状。

关键词: 大壁虎; 有鳞类; 线粒体基因组全序列; 基因结构
中图分类号: Q959.6 **文献标识码:** A **文章编号:** 0254 - 5853(2005)02 - 0123 - 06

The animal mitochondrial (mt) gene content is highly conserved, it usually encodes 13 proteins, 2 ribosomal RNAs (rRNAs), and 22 transfer RNAs (tRNAs). The gene order is also highly conserved among most vertebrates (Boore, 1999). However, gene loss and gene rearrangement have been found in some taxa (Joshua et al, 2003). Comparisons of mitochondrial systems are useful for modeling genome evolution and phylogenetic inference (Boore, 2004). 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 (Garesse et al, 1997; Boore, 2004). During the last ten years, mitochondrial

Received date: 2005 - 01 - 20; Accepted date: 2005 - 03 - 02
Foundation item: This work was supported by the Program for the National Natural Science Foundation of China(30170113; 30470249)
* Corresponding author (通讯作者), E-mail: kyzhounj@126.com
收稿日期: 2005 - 01 - 20; 接受日期: 2005 - 03 - 02
基金项目: 国家自然科学基金资助项目(30170113; 30470249)

genome sequence and gene arrangement comparisons were employed as powerful new tools for resolving ancient phylogenetic relationships (Zardoya & Meyer, 1998; Janke et al, 2001; Zhou, 2001; Joshua et al, 2003; Wu et al, 2003; Boore, 2004).

Geckos and pygopods comprise the Gekkota, one of three major lineages of living lizards and the only such lineage to be primarily nocturnal. Current estimates of gekkotan diversity recognize approximately 1 110 species in 116 genera (Kluge, 2001; Bauer, 2002) and many new taxa are described each year. Although partial sequences of *ND*1, *ND*2, *CO*1, 12*S*, tRNAs, 16*S* and Cyt *b* genes, etc. from gekkonid lizards have been published (Donnellan et al, 1999; Han et al, 2001; Macey et al, 1999; Nogales et al, 1998; Austin et al, 2004; Zhou, 2001), no complete mitochondrial sequence has been reported from any of the geckos so far. In this article, the gene sequences and organization of mitochondrial genome of tokay (*Gekko gecko*) are reported.

1 Materials and Methods

1.1 Specimen information

The specimen of *G. gecko* was obtained in Nanning, Guangxi Province in China. The fresh organs were frozen immediately in liquid nitrogen and stored under −80 ℃.

1.2 PCR amplification, cloning and sequencing

Mitochondrial DNA was extracted from liver tissue following the procedure described in Arnason et al (1991).

We obtained an mtDNA sequence of *Gekko* within *ND*4 using the primer pair (LND4: 5′-CAG TAT ATG GGC TTC ATC TC-3′; HLEU: 5′-TAC TTT TAC TTG GAT TTG CAC CA-3′) designed based on alignment of reptile mtDNAs, and a sequence of 12*S* rRNA fragment amplified by universal primers L1091 and H1478. Two pairs of primers were synthesized for long-PCR amplification based on the above two sequences: G12SL: 5′-AGG TCG AGG TGT AGC AAA CGA GAG GGA AGA GAT G-3′; GND4H: 5′-GCG ATG ATG GAT TTG AGA TCT GTT TGA CGC AGG C-3′and GND4L: 5′-ATT CAC ACA CGA ACC CTC ATC TTA ACA CGA GGC C-3′; G12SH: 5′-GCT CCT CTA GGA TGA TAT GGA ACA CCG TCA AGT C-3′. Two fragments of about 10 kb and 6 kb respectively were obtained using these two pairs of primers by long-PCR amplification (Expand Long Template PCR System, Roche).

The above two fragments were cleaved with restriction enzyme *Hin*d Ⅲ, *Eco*R Ⅰ, *Pst* Ⅰ and *Acc* Ⅰ. Fragments suitable in size were cloned using either the TA cloning method utilizing the plasmid vector pUC19, or cloning to vector pGEM-3Zf (+) by cloning kit (Takara) according to the manufacturer's protocol. The primer walking strategy was used for the fragments longer than 1 kb.

Both cloning and PCR fragments were sequenced from both ends with internal primers using ABI PRISM 310 sequencer (Perkin Elmer). All overlaps between the clones were confirmed by direct sequencing of PCR products amplified from the purified mtDNA template. The whole mitochondrial genome was read at least two times.

1.3 Data analysis

The gene sequences of *Gekko* mitochondrial genome were identified by sequence comparison with published reptile gene sequences using Clustal X 1.8 (Thompson et al, 1999). This program was also used to calculate the following data: percentage of total base substitutions; base percentages of noncoding and control region (s); transition/transversion.

Start and stop codons were used to help defining the sequences of protein-coding genes. In addition, tRNA genes were folded to verify their secondary structures using program RNAdraw and improved manually. Gene compactness, percentage of GC base component, Base skew overall, and base skew at 3rd codon position for protein-coding genes, etc. were also calculated to analyze the genome characters.

Three other mitochondrial genome sequences retrieved from GenBank were analyzed and calculated also in this study for comparing the above characters: *Eumeces egregius* (GenBank accession no. AB016606; Kumazawa & Nishida, 1999); *Iguana iguana* (GenBank accession no. AJ404872, Janke et al, 2001); *Dinodon semicarinatus* (GenBank accession no. AB008539, Kumazawa et al, 1998). Although the sequences used in this study are retrieved from defined taxonomic entities (species/subspecies), for the sake of brevity, we refer to these taxa only by their genus names in the text.

2 Results and Analyses

The complete mtDNA sequence has been determined for *G. gecko* (GenBank accession no. AY282753). The mtDNA is 16 435 bp in size, encodes 13 protein genes, 22 tRNA genes, and 2 rRNA genes. Both gene content and gene order are typical of vertebrates (Saccone et al, 1999). With respects to the length of intergenic spacers and overlaps, *Gekko* has a

rather compact genome compared with the other three squamates (Tab. 1 and Tab. 2). The overall base composition of *Gekko* mtDNA for the L-strand is: A, 32.3%; C, 27.6%; G, 13.6%; and T, 26.4%. Guanine (G) is the rarest nucleotide and GC content is 41.2%. All the *Gekko* mitochondrial protein-coding genes start with an ATG codon except *ND2* and Cyt *b* which start with ATA, and *ND3* and *ND5* which start with GTG, respectively (Tab. 1). Four protein-coding genes terminate with TAA, two end with TAG, two end with AGA, and each of the other five has an incomplete stop codon, a single stop nucleotide T, where the post-transcriptional polyadenylation can produce a standard TAA stop codon. The direction or the encoding-strand selection of the genes is identical to the typical vertebrates (Tab. 1).

The 22 tRNA genes range from 63 to 76 nt in length. Most of them could be folded into the canonical cloverleaf secondary structure (Fig. 1). The complete dihydrouridine arms (D-arms) are lacking in tRNACys and tRNASer (AGY).

The stem and loop structure of origin of the puta

Tab. 1 Organization and features of Gekko gecko mitochondrial genome

Gene/Regin	Start position	Stop position	Spacer(+) Overlap(−)	Start codon	Stop codon	Size (bp)	aa	Strand[1]
tRNA-Phe	1	73				73		H
12*S* rRNA	74	1 033	+3			960		H
tRNA-Val	1 037	1 106				70		H
16*S* rRNA	1 107	2 693				1 587		H
tRNA-Leu(UUR)	2 694	2 769				76		H
ND1	2 770	3 735	+6	ATG	TAA	966	321	H
tRNA-Ile	3 742	3 813	−1			72		H
tRNA-Gln	3 813	3 885	−1			73		L
tRNA-Met	3 885	3 954	+4			70		H
ND2	3 959	4 996	−2	ATA	TAG	1 038	345	H
tRNA-Trp	4 995	5 066	+5			72		H
tRNA-Ala	5 072	5 140				69		L
tRNA-Asn	5 141	5 213				73		L
OL	5 216	5 247	−4			32		—
tRNA-Cys	5 244	5 306				63		L
tRNA-Tyr	5 307	5 372	+4			66		L
CO Ⅰ	5 377	6 921	−7	ATG	AGA	1 545	514	H
tRNA-Ser(UCN)	6 915	6 989				75		L
tRNA-Asp	6 990	7 057				68		H
CO Ⅱ	7 058	7 750	−8	ATG	AGA	693	230	H
tRNA-Lys	7 743	7 811	+3			69		H
ATPase 8	7 815	7 974	−14	ATG	T++	160	53	H
ATPase 6	7 961	8 647	−1	ATG	TAA	687	228	H
CO Ⅲ	8 647	9 430		ATG	T++	784	261	H
tRNA-Gly	9 431	9 499				69		H
ND3	9 500	9 839		GTG	T++	340	113	H
tRNA-Arg	9 840	9 908				69		H
ND4L	9 909	10 205	−7	ATG	TAA	297	98	H
ND4	10 199	11 582	+12	ATG	T++	1 384	461	H
tRNA-His	11 571	11 642				72		H
tRNA-Ser (AGY)	11 643	11 707				65		H
tRNA-Leu (CUN)	11 708	11 778	−21			71		H
ND5	11 758	13 567	+4	GTG	T++	1 810	603	H
ND6	13 564	14 073		ATG	TAG	510	169	L
tRNA-Glu	14 074	14 149	−10			76		L
Cyt *b*	14 140	15 287		ATA	TAG	1 152	383	H
tRNA-Thr	15 288	15 357				70		H
tRNA-Pro	15 358	15 423	−2			66		L
CR	15 422	16 435				1 014		—

[1] H = heavy strand; L = light strand.

Tab. 2 Comparisons of mitochondrial genome features in four squamates

Genome character	*Gekko*	*Dinodon*	*Eumeces*	*Iguana*
Gene compactness[1]	+41; -78	+97; -33	+172; -24	+56; -22
GC base component(%)	41.2	45.2	44.2	39.9
Base skew overall	GC = -0.34; AT = 0.10	GC = -0.40; AT = 0.18	GC = -0.30; AT = 0.09	GC = -0.39; AT = 0.16
Base skew at 3rd codon position	GC = -0.52; AT = 0.22	GC = -0.78; AT = 0.48	GC = -0.56; AT = 0.36	GC = -0.80; AT = 0.57

[1] "+": Spacer; "-": Overlap. Control region(s) is excluded.

Tab. 3 Comparison of organization in noncoding region(s) among *Gekko* and 3 squamates

Taxon	Total (bp)	NC (%)	CR		ΨCR		CR Tan-rep	ΨCR Tan-rep	Reference
			Lenth (bp)	Tan-rep (%)	Lenth (bp)	Tan-rep (%)			
Gekko	16 435	6.6	1 013	36.5	—	—	2×75+22; 2×54; 2×9; 2×36	—	This study
Dinodon	17 191	12.3	1 018	18.4	1 055	17.7	2×20; 3×49	2×20; 3×49	Kumazawa et al, 1998
Eumeces	17 407	11.7	1 863	39.9	245[a]	98.0[a]	5×83+32; 2×56; 2×71+43	2×120	Kumazawa & Nishida, 1999
Iguana	16 633	7.5	1 191	9.4	—	—	2×38; 3×12	—	Janke et al, 2001

NC = Nocoding region; CR = Control region; ΨCR = Second control or noncoding region. [a] There is an 83-bp overlap between the first repeat unit (5′) and Cyt *b* gene (3′) sequence, and another 30-bp overlap between the second repeat unit (3′) and the tRNAThr gene (5′) sequence.

tive light-strand replication locates between the tRNA-Asn and tRNA-Cys. The D-loop (control region) located between tRNAPro and tRNAPhe is 1 013 bp in size.

3 Discussion

3.1 Base composition and nucleotide bias

Vertebrate animal mitochondrial genomes deviate from a random usage of nucleotide. Saccone (1999) used the formula, base-skew = (A - T/A + T) or (G - C/G + C), to evaluate the degree of the base bias, and found all the values of GC-skew were negative while all the values of AT-skew were positive in amniotes. The base bias of *Gekko* and the other three squamates (Tab. 2) accords with the feature, and we can find the base bias overall (GC = -0.34; AT = 0.10) of the *Gekko* is most similar to *Eumeces* (GC = -0.30; AT = 0.09).

The percentage of total base substitutions and gaps between *Gekko* and *Dinodon* (38.3%) is higher than that between *Gekko* and each of the other two lizards (*Gekko / Eumeces* = 33.6 %; *Gekko / Iguana* = 32.8 %), respectively; the percentage of transition/transversion between *Gekko* and *Dinodon* (73.0%) is much lower than that between *Gekko* and each of the other two lizards (*Gekko / Eumeces* = 92.2%; *Gekko / Iguana* = 84.6%).

3.2 Protein-coding genes

Base composition of vertebrate protein-coding genes is characterized by positive bias for T and negative bias for G in the second positions, but the deviations in the third positions of synonymous codons might reflect more of this feature for the light directional mutation pressure in these positions (Saccone et al, 1999). While all the amniotes have a positive bias in the third positions for C and A, most of amphibians and fishes have only a positive bias for A in this positions (Janke et al, 2001), and the base bias of *Gekko* (A = 38.8; C = 27.6; G = 8.7; T = 24.9) is similar to the later.

3.3 Transfer RNAs

The complete missing of D-arm of tRNASer (AGY) has also been found in the mtDNAs of the black spotted frog *Rana nigromaculata* (Sumida et al, 2001) and the African side-necked turtle *Pelomedusa subrufa* (Zardoya & Meyer, 1998). *Gekko* tRNA anticodon triplet sequences are exactly the same as in other vertebrates, and therefore they likely use the same genetic codons. As in other vertebrates, folding of the tRNA sequences required the formation of G + T and other atypical pairings (Fig. 1).

3.4 Noncoding regions

The stem-and-loop structure of origin for the putative light-strand replication of most vertebrates locates between tRNAAsn and tRNACys, which are absent in the African side-necked turtle (*Pelomedusa subrufa*), Chinese alligator (*Alligator sininsis*), Texas blind snake (*Typhlonectes natans*), tuatara (*Sphenodon punctatus*) and some birds (Zardoya & Meyer, 1998; Wu

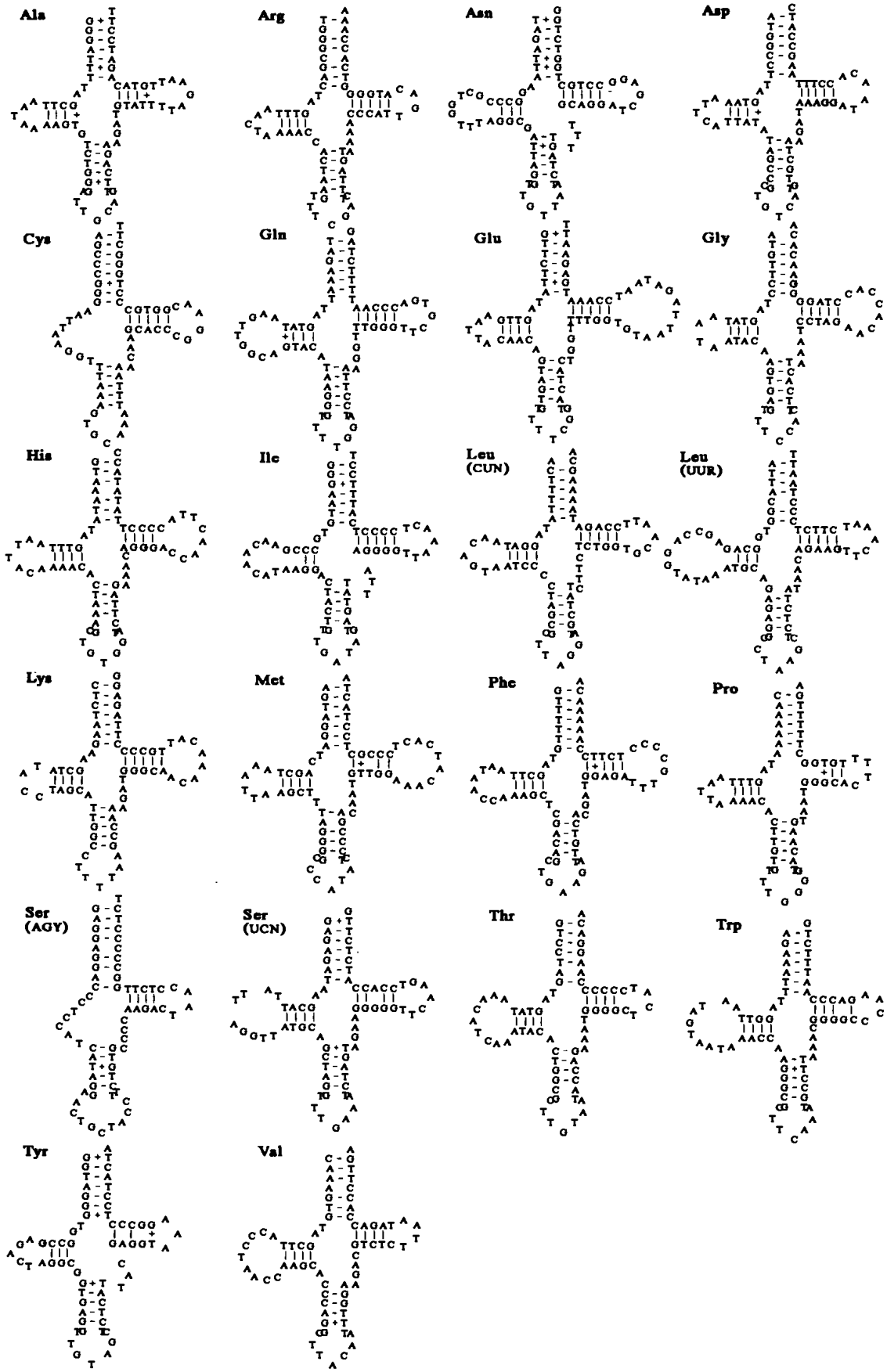

Fig. 1 Proposed secondary structures for the 22 tRNA genes of *G. gecko* mitochondrial genome

et al, 2003; Seutin et al, 1994; Kumazawa & Nishida, 1995; Desjardins & Morais, 1990), and can be easily identified at the corresponding position in *Gekko*.

The sequence of the control region (CR) located between tRNAPro and tRNAPhe is 1 013 bp in size and consists of 6.6% of the whole genome. This percentage is slightly lower than that of *Eumeces* (7.5%) and much lower than *Dinodon* (12.3%). Some structural features of the noncoding regions and their tandem-repetitive sections among the four squamates are compared in Tab. 3. The genome of *Gekko* is similar to *Iguana* in the absent of a second CR or ΨCR and differs from the latter in containing more repeat units.

Acknowledgements: We would like to thank LIU Zhong-quan, WU Xiao-bing and XU Xin-rong for kind assistance in the laboratory works and helpful suggestions on data analyses.

References:

Arnason U, Gullberg A, Widegren B. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*[J]. *J. Mol. Evol.*, **33**(6): 556 - 568.

Austin JJ, Arnold EN, Jones CG. 2004. Reconstructing an island radiation using ancient and recent DNA: The extinct and living day geckos (*Phelsuma*) of the Mascarene islands[J]. *Mol. Phylogenet. Evol.*, **31**(1): 109 - 122.

Bauer AM. 2002. Lizards[A]. In: Halliday T, Adler K. Encyclopedia of Amphibians and Reptiles[M]. Abingdon: Andromeda Oxford Ltd. 138 - 175.

Boore JL. 1999. Animal mitochondrial genomes[J]. *Nucleic Acid Res.*, **27**(8): 1767 - 1780.

Boore JL. 2004. Complete mitochondrial genome sequence of *Urechis caupo*, a representative of the phylum Echiura[J]. *BMC Genomics*, **5**(1): 67.

Desjardins P, Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome: A novel gene order in higher vertebrates[J]. *J. Mol. Biol.*, **212**(4): 599 - 634.

Donnellan SC, Hutchinson MN, Saint KM. 1999. Molecular evidence for the phylogeny of Australian gekkonoid lizards[J]. *Biol. J. Linn. Soc.*, **67**: 97 - 118.

Garesse R, Carrodeguas JA, Santiago J, Perez ML, Marco R, Vallejo CG. 1997. Artemia mitochondrial genome: Molecular biology and evolutive considerations[J]. *Comp. Biochem. Physiol. B. Biochem, Mol. Bio.*, **117**(3): 357 - 366.

Han DM, Zhou KY, Wang YQ. 2001. Phylogenetic relationships of ten gekkonid lizards (Gekkonidae: Lacertilia) from China inferred from DNA sequences of mitochondrial 12*S* ribosomal RNA gene[J]. *Acta Zool. Sin.*, **47**(2): 139 - 244. (in Chinese)

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[J]. *Proc. R. Soc. Lond. B, Biol. Sci.*, **268**(1467): 623 - 631.

Joshua SR, Jennifer CA, Christopher CA, Peter JW, Elizabeth AT, Jennifer M, David PM. 2003. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome[J]. *Mol. Phylogenet. Evol.*, **29**(2): 289 - 297.

Kluge AG. 2001. Gekkotan lizard taxonomy[J]. *Hamadryad*, **26**: 1 - 209.

Kumazawa Y, Nishida M. 1995. Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers[J]. *Mol. Biol. Evol.*, **12**(5): 759 - 772.

Kumazawa Y, Nishida M. 1999. Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: Statistical evidence for archosaurian affinity of turtles[J]. *Mol. Biol. Evol.*, **16**(6): 784 - 792.

Kumazawa Y, Ota H, Nishida M, Ozawa T. 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions[J]. *Genetics*, **150**(1): 313 - 329.

Macey JR, Wang Y, Ananjeva NB, Larson A, Papenfuss TJ. 1999. Vicariant patterns of fragmentation among gekkonid lizards of the genus *Teratoscincus* produced by the Indian collision: A molecular phylogenetic perspective and an area cladogram for Central Asia[J]. *Mol. Phylogenet. Evol.*, **12**(3): 320 - 332.

Nogales M, Lopez M, Jimenez-Asensio J, Larruga JM, Hernandez M, Gonzalez P. 1998. Evolution and biogeography of the genus *Tarentola* (Sauria: Gekkonidae) in the Canary Islands, inferred from mitochondrial DNA sequences[J]. *J. Evol. Biol.*, **11**(4): 481 - 494.

Saccone S, Giorgi CD, Gissi C. 1999. Evolutionary genomics in Metazoa: The mitochondrial DNA as a model system[J]. *Gene*, **238**(1): 195 - 209.

Seutin G, Lang BF, Mindell DP, Morais R. 1994. Evolution of the WANCY region in amniote mitodhondrial DNA[J]. *Mol. Viol. Evol.*, **11**(3): 329 - 340.

Sumida M, Kanamori Y, Kaneda H, Kato Y, Nishioka M, Hasegawa M, Yonekawa H. 2001. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*[J]. *Genes Genet. Syst.*, **76**(5): 311 - 325.

Thompson JD, Higgins DG, Gibson TJ. 1999. CLUSTAL X Multiple Sequence Alignment Program, version 1.8.[CP]

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[J]. *Chinese Science Bulletin*, **48**(18): 1954 - 1958. (in Chinese)

Zardoya R, Meyer A. 1998. Complete mitochondrial genome suggests diapsid affinities of turtles[J]. *Proc Natl Acad Sci, USA*, **95**(24): 14226 - 14231.

Zhou KY. 2001. Molecular phylogenetics of amphibians and reptiles[J]. *Zooligical Research*, **22**(5): 397 - 405. (in Chinese)