

蒙古狼线粒体基因组序列分析及其系统发育地位

陈磊^{1,2}, 张洪海^{2,*}, 马建章¹

(1. 东北林业大学野生动物资源学院, 哈尔滨 150040; 2. 曲阜师范大学生命科学学院, 曲阜 273165)

摘要:应用 Long-PCR 和克隆测序法得到蒙古狼(*Canis lupus chanco*)线粒体基因组全序列, 结合 GenBank 中现有犬科动物线粒体基因组数据, 应用最大简约法(MP)、最大似然法(ML)和 Bayesian 分析法对蒙古狼的系统发育地位进行了探讨。结果如下: 蒙古狼线粒体基因组全长 16709 bp, 包含 13 个蛋白质编码基因、22 个 tRNA 基因、2 个 rRNA 基因和 1 个非编码区。序列碱基的组成存在明显的 A-T 偏好性。tRNA 基因中除 tRNA-Ser(AGY)缺少双氢尿嘧啶(DHU)臂以外, 其余均能折叠成典型的三叶草二级结构。大多数蛋白质编码基因的起始和终止密码子与犬科动物有报道相同, COX II 基因的起始密码子为 ATA, 与其他犬科动物不同。基于 12S rRNA + 16S rRNA + H 链上的 12 个蛋白质编码基因的联合数据的系统发育分析发现, 在已报道的狼亚种数据中, 西藏狼(*Canis lupus laniger*)的分化时间最早, 其次为阿拉伯狼(*Canis lupus arabs*), 蒙古狼与欧亚狼(*Canis lupus lupus*)的系统发育地位最为接近。

关键词:线粒体; 基因组; 蒙古狼; 系统发育

The mitochondrial genome of the Mongolian wolf *Canis lupus chanco* and a phylogenetic analysis of *Canis*

CHEN Lei^{1, 2}, ZHANG Honghai^{2,*}, MA Jianzhang¹

1 College of Wildlife Resource, Northeast Forestry University, Harbin 150040, China

2 College of Life Science, Qufu Normal University, Qufu 273165, China

Abstract: The complete mitochondrial genome sequence of the Mongolian wolf, *Canis lupus chanco*, was determined by the long and accurate polymerase chain reaction (L-PCR) based on the PCR fragments of primers. In order to clarify the phylogenetic position of *Canis lupus chanco*, we constructed the phylogenetic tree based on the concatenated 12S rRNA + 16S rRNA + 12 heavy-strand protein-coding genes of 12 Canidae animals with the genomic data obtained from GenBank. Here are the results: The entire mtDNA sequence is 16709 base pairs (bp) in length and contains two ribosomal RNA genes, 13 protein-coding genes, 22 transfer RNA genes and one control region. A strong A-T bias was found in the nucleotide sequence of mitochondrial genome of the Mongolian wolf. All the transfer RNA genes could be folded into the typical cloverleaf-shaped structure except tRNA-Ser (AGY), which lacked the "DHU" arm. Most of the protein-coding genes have the same codon usage as other Canidae animals except COX II gene, which terminates with the stop codon ATA. The phylogenetic analysis indicated that within the subspecies data of *Canis lupus*, *Canis lupus laniger* polarized first, then did *Canis lupus arabs*. *Canis lupus chanco* and *Canis lupus lupus* were sister species, which had similar phylogenetic positions.

Key Words: mitochondrial genome; *Canis lupus chanco*; phylogenetic relationship

由蛋白质编码基因、tRNA 基因、rRNA 基因和非编码区构成的线粒体基因组在不同脊椎动物物种间呈高

基金项目:教育部 2007 年新世纪优秀人才支持计划资助项目(NCET-07-0507);山东省自然科学基金重点资助项目(Z2008D01);山东省中青年科学家奖励基金资助项目(2005BS02005);山东省科技攻关计划资助项目(2007GG2009011)

收稿日期:2009-08-17; **修订日期:**2009-12-01

* 通讯作者 Corresponding author. E-mail: zhanghonghai67@126.com

度保守状态^[1],该基因组蕴含着分子动力学、比较基因组学和物种进化研究所需的大量遗传信息^[2]。自从从人类的线粒体基因组被测序以来^[3],大量的脊椎动物线粒体基因组信息不断涌现^[4-5]。这些线粒体基因组信息的出现促进了动物分子进化领域研究的发展^[6],人们开始通过比较线粒体基因组的某些基因来推断物种间的系统进化关系^[7]。由于通过比较不同基因所得出的结论往往不尽相同^[8-10],进行线粒体基因组全序列的研究比较逐渐成为阐述物种间的系统发育关系的有效手段^[11-13]。已有学者通过比较线粒体基因组编码区序列构建了食肉目犬型亚目(*Caniformia*, dog-like carnivorans)动物的系统发育模型^[14],其中就包括常见的犬科动物。

犬科动物作为世界上分布最广泛的哺乳动物类群之一,具有与其它哺乳动物相似的线粒体基因组构造^[15]。截至2009年8月,GenBank数据库共收录了2个属4种犬科动物的线粒体基因组数据,分别是狼(*Canis lupus*)、家犬(*Canis lupus familiaris*)、郊狼(*Canis latrans*)和赤狐(*Vulpes vulpes*)^[16]。作为犬科动物中体型最大者,狼因其广谱的生境适应性而具有非常广泛的分布^[17]。然而,不同地域种群的狼的亚种分化情况学术界尚无统一的标准。目前比较认可的是世界上现存的狼至少可划分为17个亚种^[18-19]。根据地区隔离特点,狼在中国的分布大致可以划分为5个种群:东北、华北、蒙新、青藏和西南^[20]。目前上述种群中仅有学者对分布于青海省的西藏狼(*Canis lupus laniger*)的线粒体基因组进行了研究^[21],未见其余4个地域种群的狼线粒体基因组的相关报道。

分布于中国内蒙古地区的狼在分类地位上隶属于蒙古狼亚种(*Canis lupus chanco*)^[18]。本文首次测定了蒙古狼的线粒体基因组全序列,并与GenBank中现有的9组来自不同国家和地区的狼的线粒体基因组数据相结合,分析狼的线粒体基因组结构特征。同时通过对这些狼的线粒体基因组进行比较,尝试更好的阐述它们的系统进化关系。

1 材料和方法

1.1 基因组 DNA 的提取

实验动物为一只健康成年雌性蒙古狼,捕捉于内蒙古自治区达赉湖自然保护区附近。采集新鲜血液200 μ L,总DNA的提取参照《分子克隆实验指南》^[22]进行。所得DNA溶液放置在-20℃保存。

1.2 引物设计、Long-PCR 和 Sub-PCR 扩增

根据已公布的犬科动物线粒体基因组全序列借助Oligo6.71软件^[23]辅以手工设计引物,共设计5对Long-PCR引物(表1)。

表1 蒙古狼线粒体基因组全序列所用L-PCR引物

Table 1 Primers designed for amplifying and sequencing the mitochondrial genome of the Mongolian wolf in the L-PCR

编号 Primer No.	产物长度 Product length (bp)	上游引物 Upstream primer sequence (5'→3')	下游引物 Downstream primer sequence (5'→3')
1-F/R	3275	TCCCTCTAGAGGAGCCTGTTTC	GGGTATGGGCCCCGATAGCTT
2-F/R	3311	GCGGATAAAAAGACTTACTTTGATAGAG	GCGAATTTAACTTTGACAAAGTCATGT
3-F/R	3337	GAAGAAAGGAAGGAATCGAACC	GCGTAGGGATGATAATTTTAGCATT
4-F/R	3904	GTATTTGCTGCCTGCCAAGC	TAGTGGTGGGATTGCTTCTGC
5-F/R	3534	GGGTATTGCTCAGTAGCCATAGC	GGTTTGCTGAAGATGCGGATATAT

L-PCR反应体系为50 μ L:模板DNA 1 μ L,10 $\times$ L-PCR Buffer 5 μ L,dNTP Mixture (各2.5 mmol/L) 4 μ L,上下游引物(10 pmol/L) 2 μ L,L-PCR DNA Polymerase(5 U/ μ L) 1 μ L,双蒸水补足。

L-PCR循环参数:起始变性95℃,5 min;95℃变性40 s;57℃退火30 s;72℃延伸4分30 s;35个循环;72℃延伸10 min,4℃保存。

以L-PCR产物为模板,根据已公布的欧亚狼和家犬的线粒体基因组全序列设计11对引物进行Sub-PCR(表2)。Sub-PCR反应体系与L-PCR相同,循环参数如下:起始变性95℃,5 min;95℃变性40 s;57℃退火30

s;72℃延伸 2 min 30 s;35 个循环;72℃延伸 10 min;4℃保存。

为防止污染,所有 PCR 反应均设立空白对照。

1.3 克隆及测序

将 Sub-PCR 扩增产物连接 pGEM-T Easy 载体(Promega 公司),热转化宿主大肠杆菌,蓝白斑筛选,挑取阳性菌斑应用 ABI3730 测序仪进行双向测序。

表 2 蒙古狼线粒体基因组全序列所用 Sub-PCR 引物

Table 2 Primers designed for amplifying and sequencing the mitochondrial genome of the Mongolian wolf in the Sub-PCR			
编号 Primer No.	产物长度 Product length (bp)	上游引物 Upstream primer sequence (5'→3')	下游引物 Downstream primer sequence (5'→3')
1-F/R	1654	TCCCTCTAGAGGAGCCTGTTC	TCCGAGGTCACCCCAACC
2-F/R	1740	GACGAGAAGACCCTATGGAGC	GGGTATGGGCCCGATAGCTT
3-F/R	1727	GGCGGATAAAAGAGTTACTTTGATAGAG	GCCTACTATACCGGCTCATGC
4-F/R	1593	GCTCAGCCATTTTACCTATGTTC	GCGAATTTAACTTTGACAAAGTCATGT
5-F/R	1869	GAAGAAAGGAAGGAATCGAACC	GCGAAGAGTTGTAGTGAAATCATAT
7-F/R	1591	GCTACCTAATGACCCACCAAAC	GCCTAGGGATGATAATTTTATGCAAT
8-F/R	1673	GTATTTGCTGCCTGCGAAGC	CGCTTATCTGGAGTTGCACC
9-F/R	2348	CCGCAAGAACTGCTAATTCATG	TAGTGCTGGGATTGTTGTGCG
10-F/R	1570	GGGTATTGCTCACTAGCCATAGC	GCAGAATTTACGCTTTGGGTG
11-F/R	2087	CGCGATGAAGAGTCTTTGTAGTAT	GGTTTGCTGAAGATGCGGCTATAT

1.4 序列分析

测序所得序列峰图应用 Chromas 2.23 软件(<http://www.technelysium.com.au/chromas.html>)进行分析,通过 Blast 与 GenBank 中相似序列比对后,应用 DNAMAN 6.0(Lynnon Biosoft, Quebec, Canada)辅助手工进行序列拼接。通过 ClustalX 1.83^[24]与已报道的犬科动物线粒体基因组序列进行比对确定编码蛋白质序列的准确起始位置,应用 Sequin 9.20 确定 13 个蛋白质基因的位置,应用 tRNA Scan-SE1.21(<http://lowelab.ucsc.edu/tRNAscan-SE/>)定位 21 个 tRNA 基因,通过比对确定剩余一个 tRNA-Ser(AGC)和两个 rRNA 的位置。tRNA 二级结构应用 RNA structure 4.6^[25]进行预测。应用 MEGA4.1^[26]统计碱基组成和密码子使用频率。蒙古狼线粒体基因组完全注释序列已提交 GenBank (Accession No. GQ374438)。

1.5 系统发育分析

从 GenBank 上下载 9 组不同国家和地区的狼(*Canis lupus*)以及家犬(*Canis lupus familiaris*)和赤狐(*Vulpes vulpes*)的线粒体基因组全序列(表 3),不同国家和地区的狼以(拉丁学名+地区名)进行命名(例如:源于加拿大的狼的线粒体基因组序列命名为*Canis lupus Canada*)。应用 12S rRNA+16S rRNA+H 链上的 12 个蛋白质编码基因进行系统发育树的构建。L 链上的 ND6 基因由于在核苷酸和氨基酸组成上与其他蛋白质编码基因有明显差异而在构建系统发育矩阵时被排除。

系统发育分析以赤狐(*Vulpes vulpes*)为外群,采用最大简约法(maximum parsimony, MP)、最大似然法(maximum likelihood, ML)和贝叶斯法(Bayesian infer-

表 3 本研究中所用到的线粒体基因组序列

Table 3 The taxa and sequence accession number of the twelve species used for phylogenetic analysis

物种 Species	GenBank 检索号 GenBank accession number
<i>Canis lupus</i> Canada	DQ480508
<i>Canis lupus</i> Sweden 1	AM711902
<i>Canis lupus</i> Sweden 2	DQ480504
<i>Canis lupus</i> Tibetan 1	FJ032363
<i>Canis lupus</i> Tibetan 2	EU442884
<i>Canis lupus</i> Spain	DQ480505
<i>Canis lupus</i> Russia	DQ480503
<i>Canis lupus</i> Saudi Arabia 1	DQ480506
<i>Canis lupus</i> Saudi Arabia 2	DQ480507
<i>Canis lupus familiaris</i>	CFU96639
<i>Vulpes vulpes</i>	NC_008434
<i>Canis lupus mongolian</i>	GQ374438(本文)

ence, BI) 进行。

MP 和 ML 分析应用 PAUP* 4b10 软件包完成^[27-28]。选择树二等分再连接 (tree-bisection reconnection, TBR) 交换算法和启发式搜索, 100 次随机加入序列, multrees 参数设置为默认, 应用非参数自展法 1000 次重复 (replicates = 1000) 检验各节点的支持率^[29-30]。相等树的非强制性拓扑结构树形用 Kishino-Hasegawa 似然性检测来比较, 最佳拓扑结构 (最小 $-\ln L$) 被选用。所有位点均作无序处理, 分析前排除所有无信息位点。

BI 分析应用 MrBayes3.1.2 软件完成^[31]。根据 MODELTEST v. 3.7 软件中等级似然率 (hLRT) 和 AIC 检验标准来选择替代模型, 最佳替代模型为: GTR + G + I (a general time reversible + gamma + invariant)^[32-33]。以后验概率 (posterior probability, PP) 表示各分支的可信度^[34], 以 PP > 95% 为有效支持^[35]。起始树设为随机树, 马尔科夫链的蒙特卡洛方法设置为 4 条链 (3 条 heat 链和 1 条 cold 链) 同时运行 2×10^6 代。每 100 代对系统树进行抽样, 最终得到 20001 棵系统发育树, 重复一次以确保 MCMC 的收敛。将运行过程中所得的冷链对数似然值 ($-\log$ likelihood scores, $-\ln L$) 与相应的代数进行作图, 找到对数似然值达到饱和的位置, 到达饱和前的数据被作为老化样本而舍弃。在舍弃老化样本后, 共有 15001 个样本用于构建一致树和计算相关参数。

2 结果

2.1 蒙古狼线粒体基因组组成结构

蒙古狼线粒体基因组全长 16709 bp, 为环形双链 DNA, 包括 13 个蛋白质编码基因、22 个 tRNA 基因、2 个 rRNA 基因 (12S rRNA 和 16S rRNA) 和 1 个非编码区即控制区。除 ND6 等 9 个基因位于 L 链上外, 其余基因全部由 H 链编码 (图 1)。蒙古狼线粒体基因组碱基组成存在明显的 A-T 偏好性, 总的碱基顺序为 A (31.6%) > T (28.7%) > C (25.5%) > G (14.2%)。A-T 碱基对含量明显高于 G-C 碱基对含量。

2.1.1 蛋白质编码区

蒙古狼的蛋白质编码基因包括 7 个 NADH 脱氢酶亚基基因 (ND1、ND2、ND3、ND4L、ND4、ND5 和 ND6), 3 个细胞色素氧化酶亚基基因 (COX I、COX II、COX III), 2 个 ATP 酶亚基基因和 1 个细胞色素 *b* 基因, 总长度为 11418 bp。这些蛋白质编码基因均不包含内含子, 有些基因与相邻基因存在少数碱基的间隔或重叠, 其中以 ATP8 和 ATP6 酶亚基基因的碱基重叠最多 (43bp)。

在密码子的使用方面, 除 ND2、ND3、ND5 和 COX II 以 ATA 为起始密码子外, 其余均以 ATG 作为起始密码子。终止密码子以 TAA 为主, ND2 基因为 TAG, 细胞色素 *b* 基因为 AGA, COX III、ND3 和 ND4 基因均为不完整的 T (或 TA) (表 4)。

2.1.2 tRNA 基因和 rRNA 基因

蒙古狼线粒体基因组包含 22 个 tRNA 基因 (表 5), 其长度为 60bp (tRNA-Ser) — 75bp (tRNA-Leu)。除 tRNA-Gln、tRNA-Tyr、tRNA-Ala、tRNA-Asn、tRNA-Cys、tRNA-Glu、tRNA-Pro 和 tRNA-Ser (UCN) 由 L 链编码以外, 其余 tRNA 基因均由 H 链编码。在这些 tRNA 基因中, 除 tRNA-Ser (AGY) 缺少双氢尿嘧啶 (DHU) 臂以外, 其余 21 中 tRNA 均能形成典型的三叶草二级结构 (图 2)。

蒙古狼线粒体基因组包含 2 个 rRNA 基因, 即 12S rRNA 和 16S rRNA, 长度分别为 954bp 和 1581bp, 其位置与其他犬科动物相同。这两个基因的 A + T 含量分别为 62.2% 和 61.6%, 明显高于各自的 G + C 含量

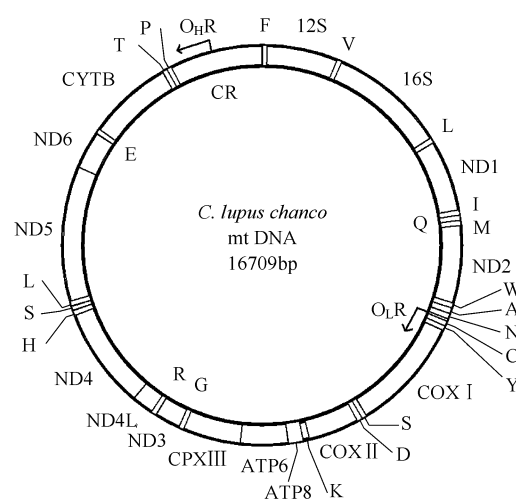


图 1 蒙古狼线粒体基因组基因构成

Fig. 1 The organization of the complete mitochondrial genome of *Canis lupus chanco*

tRNA 基因以相应氨基酸的单个大写字母表示; 其他基因的缩写分别是: ND1—6 为 NADH 脱氢酶亚基 1—6; COX I—III 为细胞色素氧化酶亚基 I—II; ATP6, 8 为三磷酸腺苷酶亚基 6, 8; CYTB 为细胞色素 *b*; CR 为控制区; OH-R 和 OL-R 分别代表 H 链和 L 链的复制起点

(41.5% 和 38.4%)。

表 4 蒙古狼线粒体基因组编码区碱基组成和密码子使用情况

Table 4 The characteristics of the Mongolian wolf mitochondrial genome protein-coding genes											
基因 Genes	位置 Positions		长度/bp Size/bp	编码链 Strand	间隔(+)/重叠(-) Interval(+)/Overlap(-)	密码子 Codons		碱基组成 Base composition/%			
	从 From	到 To				Start	Stop	A	C	G	T
ND1	2748	3703	956	H	+ 2	ATG	TAA	29.7	29.3	11.8	29.2
ND2	3915	4956	1042	H		ATA	TAG	35.4	27.4	9.7	28.0
COX I	5350	6894	1545	H	+ 1	ATG	TAA	26.8	22.9	17.9	32.4
COX II	7025	7717	693	H	- 10	ATA	TAA	32.9	25.5	13.6	28.0
ATP8	7803	8006	204	H	+ 1	ATG	TAA	34.3	27.9	8.3	29.4
ATP6	7964	8644	681	H	- 43	ATG	TAA	29.1	28.3	12.2	30.4
COX III	8644	9427	784	H	- 1	ATG	TA -	25.9	27.2	16.3	30.6
ND3	9496	9842	347	H		ATA	TA -	29.7	28.0	14.1	28.2
ND4L	9913	10209	297	H		ATG	TAA	31.1	22.1	12.7	34.1
ND4	10203	11580	1378	H	- 7	ATG	T - -	31.2	27.8	10.8	30.2
ND5	11780	13600	1821	H		ATA	TAA	30.8	27.7	11.8	29.7
ND6	13584	14111	528	L	- 17	ATG	TAA	39.5	28.0	10.5	22.0
CYT B	14185	15324	1140	H	+ 4	ATG	AGA	28.9	28.0	13.9	29.1

表 5 蒙古狼线粒体基因组 tRNA 基因和 rRNA 基因组成

Table 5 The characteristics of the Mongolian wolf mitochondrial genome tRNA and rRNA genes									
基因 Genes	位置 Positions		长度/bp Size/bp	编码链 Strand	基因 Genes	位置 Positions		长度/bp Size/bp	编码链 Strand
	从 From	到 To				从 From	到 To		
tRNA-Phe	1	69	69	H	tRNA-Tyr	5281	5348	68	L
12S rRNA	70	1023	954	H	tRNA-Ser(UCN)	6892	6960	69	L
tRNA-Val	1024	1090	67	H	tRNA-Asp	6967	7034	68	H
16S rRNA	1090	2670	1581	H	tRNA-Lys	7735	7801	67	H
tRNA-Leu(UUR)	2671	2745	75	H	tRNA-Gly	9428	9495	68	H
tRNA-Ile	3704	3772	69	H	tRNA-Arg	9843	9912	70	H
tRNA-Gln	3770	3843	74	L	tRNA-His	11581	11649	69	H
tRNA-Met	3845	3914	70	H	tRNA-Ser(AGY)	11650	11709	60	H
tRNA-Trp	4957	5024	68	H	tRNA-Leu(TAG)	11710	11779	70	H
tRNA-Ala	5038	5106	69	L	tRNA-Glu	14112	14180	69	L
tRNA-Asn	5108	5179	71	L	tRNA-Thr	15325	15394	70	H
tRNA-Cys	5214	5280	67	L	tRNA-Pro	15394	15459	66	L

2.1.3 非编码区

蒙古狼的线粒体基因组控制区长度为 1250 bp,位于 tRNA-Pro 和 tRNA-Phe 之间,其碱基组成为:A: 27.5%、C:26.6%、G:16.9%、T:29.0%。控制区 3 个部分的长度分别为:延长终止相关序列(ETAS):222 bp,中央保守区域(CD):374 bp,保守序列区(CSB):654 bp。

蒙古狼线粒体基因组 L 链的复制起点(O_L)位于 WANCY 基因簇(tRNA-Trp、tRNA-Ala、tRNA-Asn、tRNA-Cys、tRNA-Tyr)之间,长度为 34 bp。

2.2 系统发育分析

应用最大简约法分析得出 MP 树树长为 2435,CI=0.955,RI=0.833。图 3(A)显示了 50% 主要一致树。分析发现,犬属动物明显分为两大支系:西藏狼作为一个亚种,单独构成第一分支(BP=100%),第二分支中,阿拉伯狼位于该分支基部(BP=100%);瑞典狼、西班牙狼和加拿大狼构成一个支系(BP=92%),这一支系与蒙古狼成姊妹群关系(BP=99%),再与俄罗斯狼(BP=53%)和家犬(BP=61%)聚在一起。

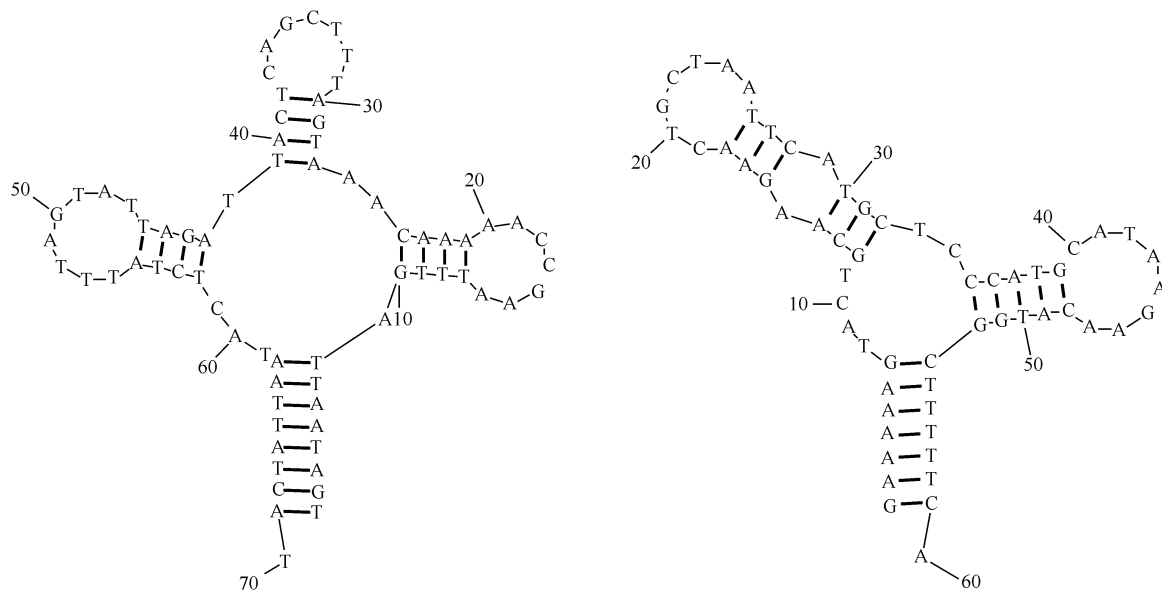


图2 tRNA 二级结构图

Fig. 2 The secondary structure of tRNA

(A) :tRNA-Arg 二级结构:呈典型的三叶草形,由氨基酸接收臂,TΨC 臂、反密码子臂和双氢尿嘧啶(DHU)臂组成;(B) :tRNA- Ser(AGY) 二级结构:缺乏双氢尿嘧啶(DHU)臂

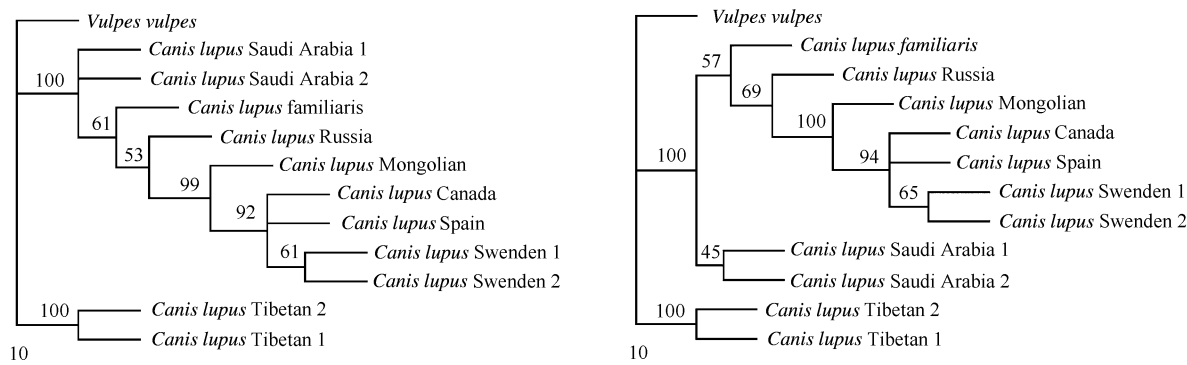


图3 以赤狐为外群应用 PAUP4.0 * b10 软件最大简约法构建的 MP 树(A)和最大似然法构建的 ML 树(B) (Bootstrap 50% majority-rule consensus tree)

Fig. 3 Phylogenetic analysis of selected *Canis* inferred from concatenated of the nucleotide sequences of 12S rRNA + 16S rRNA + 12 heavy-strand protein-coding genes using maximum parsimony (MP) (A) and maximum likelihood (ML) (B) methods

该树系应用 TBR 交换算法基于 12S rRNA + 16S rRNA + 12 个 H 链蛋白质编码基因的联合数据进行 100 次随机抽样构建;MP 树长为 2435, CI=0.955, RI=0.833;左下角图例表示枝长值;节点处的数值表示 1000 次非参数自展检验 (Number of bootstrap replicates = 1000) 所得支持率值

最大似然法分析得出的 ML 树和贝叶斯法分析所得的 Bayesian 树与 MP 树的拓扑结构十分接近。这两种方法的分析结果同样支持西藏狼作为单独分支最先分化出来 (ML: BP = 100; BI: PP = 1.0), 另一分支中瑞典狼、加拿大狼和西班牙狼构成同一支系的自展支持率为 94%, 后验概率为 100%。蒙古狼与这 3 种狼的姊妹群关系也得到了很大支持 (ML: BP = 100; BI: PP = 1.0)。最大似然法和贝叶斯法的分析结果更支持阿拉伯狼作为一个单独的支系存在 (ML: BP = 100; BI: PP = 1.0), 这一点与最大简约法分析结果略有差别。

3 讨论

目前已知最大的脊椎动物的线粒体基因组是鱼纲的宽咽鱼 (*Eurypharynx pelecyanoides*, 18978bp)^[36]。哺乳

动物线粒体基因组长度一般为 16—17 kb。已报道的犬科动物中线粒体基因组最长的是赤狐 (*Vulpes vulpes*, 16813 bp)。不同地区的狼和家犬的线粒体基因组全序列长度都集中在 16700—16800 bp 之间,蒙古狼线粒体基因组序列的长度是目前已知的犬科动物中最短的,但其长度与已报道的瑞典狼 (GenBank Accession No. DQ480504) 相差仅有 1 bp,其基因组成和排列顺序与其它犬科动物完全相同。脊椎动物线粒体基因组的碱基组成具有明显的 A-T 偏好性,这一点在犬科动物中也有体现。已报道的犬科动物的线粒体基因组序列中,A-T 碱基对的含量除赤狐最低为 59.1% 以外,其余均在 60.3%—60.5% 之间。这些特征均说明犬科动物线粒体基因组在进化过程中的高度保守性。

动物线粒体基因组中各蛋白质编码基因的起始和终止密码子同样具有高度的保守性,起始密码子多为 ATG,终止密码子主要是 TAA^[37]。蒙古狼线粒体基因组编码区基因符合这一规律。两栖类^[38-39]、爬行类^[40-42]、鸟类^[43]和线虫类^[44]的线粒体基因组编码区 *cyt b* 基因主要以 T、TA 或 TAA 终止,而哺乳动物线粒

体基因组 *cyt b* 基因多以 AGA 为终止密码子,这在家猪^[45]、家兔^[46]、猫科动物^[47]和诸多犬科动物^[15-16,21]都有体现。蒙古狼线粒体基因组中 *cyt b* 基因终止密码子是 AGA,符合哺乳动物的一般特征。

与已报道的犬科动物相同,蒙古狼线粒体编码区 ND2 基因的终止密码子为 TAG, COX III、ND3 和 ND4 基因终止密码子均为不完整的 T (或 TA)。所不同的是,西藏狼、欧亚狼和家犬的 COX II 基因多以 ATG 起始,而蒙古狼 COX II 的起始密码子为 ATA;蒙古狼 COX II 基因终止密码子为 TAA,与欧亚狼和家犬相同,不同于西藏狼 (TAG)^[15-16,21]。

一般认为,脊椎动物的线粒体 tRNA-Met, tRNA-His, tRNA-Leu (CUN) 等序列保守性最高,可达 80% 以上, tRNA-Ser (AGY) 变化最大,同源率仅为 54%—64%。tRNA 基因最保守区域为反密码子环,其次为茎区,变异最大的为 DHU 臂和 T ψ C 臂。通过对雪豹等猫科动物和多种犬科动物的研究发现,哺乳动物的线粒体基因组中 tRNA-Ser (AGY) 往往缺少 DHU 臂^[47-48],这一点在蒙古狼线粒体基因组也有体现。蒙古狼的 22 个 tRNA 基因的排列顺序、碱基组成、长度和二级结构与所报道的犬科动物的研究结果高度一致。

所构建的 MP 树、ML 树和 Bayesian 树均显示瑞典狼、西班牙狼和加拿大狼明显聚为一支,可见 3 个地区的狼从分类地位上应该都是欧亚狼 (*Canis lupus lupus*) 亚种。3 种进化树均显示蒙古狼与欧亚狼成姊妹群关系,且得到了很高的自展率 and 后验概率支持,证明蒙古狼与欧亚狼具有相似的分化时间和系统发育地位。由于狼具有强大的迁徙能力,不同地理种群的狼之间可能存在着基因交流,所以以地区为单位的狼的群体不一定具有独立的遗传结构^[20,49]。由于人们对世界各地狼的迁徙过程所知甚少,加上现有线粒体基因组数据的局限性,目前尚不能对蒙古狼与欧亚狼系统发育关系较近的原因做出解释。

有学者认为,分布于中国西藏地区的狼可能是世界上最古老的狼的亚种^[12,50]。通过应用线粒体基因组控制区进行系统发育研究发现,中国青藏群体、西南群体的狼与东北群体、华北群体是分属于两个不同进化时间的类群,可作为独立亚种进行划分^[20]。通过应用最大简约法、最大似然法和 Bayesian 分析均证明,西藏狼 (*Canis lupus laniger*) 作为一个亚种,其分化时间最早,之后是分布于阿拉伯地区的阿拉伯狼 (*Canis lupus arabs*)。这一点为前人的推测提供了支持^[12,50]。另有学者通过运用线粒体基因组进行犬科动物系统发育关

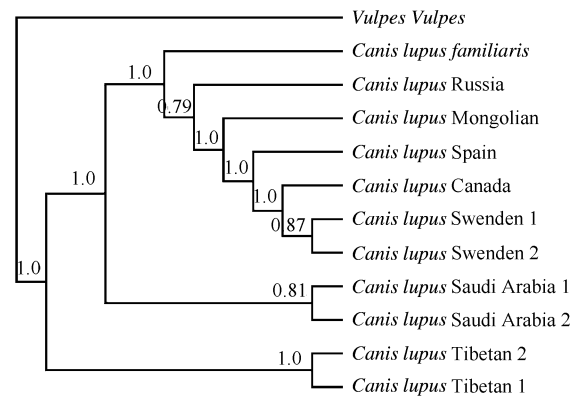


图4 以赤狐为外群应用 MrBayes3.1.2 软件通过 Bayesian 分析构建的贝叶斯树

Fig. 4 Phylogenetic analysis of selected *Canis* inferred from concatenated of the nucleotide sequences of 12S rRNA + 16S rRNA + 12 heavy-strand protein-coding genes using Bayesian (BI) methods

该树系应用 GTR + G + I 进化模型基于 12S rRNA + 16S rRNA + 12 个 H 链蛋白质编码基因的联合数据构建;节点处上方的数值表示该节点的后验概率 (posterior probability)

系构建和遗传距离推算,其结果不支持将狼和家犬作为两个种进行划分^[14]。本文研究结果也发现,家犬与蒙古狼(*Canis lupus chanco*)和欧亚狼(*Canis lupus lupus*)的分化时间要晚于以上两种狼与西藏狼(*Canis lupus laniger*)的分化时间。西藏狼(*Canis lupus laniger*)作为较早分化的一个亚种,除了说明其独特的遗传背景之外,是否还说明其分化时间要早于家犬的家养驯化时间,这一推测尚需大量考古学、形态学和进化生物学领域后续实验来证明。

References:

- [1] Boore J L. Animal mitochondrial genomes. *Nucleic Acids Research*, 1999, 27(8):1767-1780.
- [2] O'Brien S J. Genetic and phylogenetic analyses of endangered species. *Annual Review of Genetics*, 1994, 28:467-489.
- [3] Anderson S, Bankier A T, Barrell B G, Bruijn M H L de, Coulson A R, Drouln J, Eperon I C, Nierlich D P, Roe B A, Sanger F, Schreier P H, Smith A J H, Staden R, Young I G. Sequence and organization of the human mitochondrial genome. *Nature (London)*, 1981, 290(9):457-465.
- [4] D'Erchia A M, Gissi C, Pesole G, Saccone C, Arnason U. The guinea-pig is not a rodent. *Nature*, 1996, 381:597-600.
- [5] Penny D, Hasegawa M. The platypus put in its place. *Nature*, 1997, 387:549-550.
- [6] Arnason U. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99 (12):8151-8156.
- [7] Kumazawa Y, Nishida M. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *Journal of Molecular Evolution*, 1993, 37:380-398.
- [8] Flynn J J, Nedbal M A. Phylogeny of the Carnivora (Mammalia): congruence vs incompatibility among multiple data sets. *Molecular Phylogenetics and Evolution*, 1998, 9: 414-426.
- [9] Flynn J J, Finarelli J A, Zehr S, Hsu J, Nedbal M A. Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. *Systematic Biology*, 2005, 54 (2): 317-337.
- [10] Cao Y, Adachi J, Janke A, Paabo S, Hasegawa M. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of tree based on a single gene. *Journal of Molecular Evolution*, 1994, 39:519-527.
- [11] Boore J L, Macey J R, Medina M. Sequencing and comparing whole mitochondrial genomes of animals. *Methods in Enzymology*, 2005, 395:311-348.
- [12] Savolainen P, Zhang Y P, Luo J, Lundeberg J, Leitner T. Genetic evidence for an east Asian origin of domestic dogs. *Science*, 2002, 298 (5598):1610-1613.
- [13] Leonard J A, Wayne R K, Wheeler J, Valadez R, Guillén S, Vila C. Ancient DNA evidence for old world origin of new world dogs. *Science*, 2002, 298(5598):1613-1616.
- [14] Arnason U, Gullberg A, Janke A, Kullberg M. Mitogenomic analyses of caniform relationships. *Molecular Phylogenetics and Evolution*, 2007, 45: 863-874.
- [15] Kim K S, Lee S E, Jeong H W, Ha J H. The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. *Molecular Phylogenetics and Evolution*, 1998, 10:210-220.
- [16] Delisle I, Strobeck C. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. *Molecular Phylogenetics and Evolution*, 2005, 37: 192-201.
- [17] Harrington F H, Paquet P C. *Wolves of the World, Perspectives of Behavior, Ecology and Conservation*. New Jersey, USA: Noyes Pub, Park Ridge. 1982. 474-545.
- [18] Mech L D. *Mammalian species: Canis lupus*. USA: The american society of mammalogists. No. 37. 1974: 1-6.
- [19] Mech L D. *Wolves: Behavior, Ecology, and Conservation*. USA: University of Chicago Press, 2003: 428.
- [20] Luo L Y, JIN L, Yang G S. The phylogenetic relationship and subspecies differentiation of Chinese wolf. *Journal of Northeast Forestry University*, 2003, 31(5): 81-83.
- [21] Meng C, Zhang H H, Chen Y C. Sequencing and analysis of mitochondrial genome of Chinese grey wolf. *Chinese Journal of Biochemistry and Molecular Biology*, 2008, 24(12): 1170-1176.
- [22] Sambrook J, Russell D W. *Molecular Cloning: A Laboratory Manual*, 3 rd ed. New York: Cold Spring Harbor Laboratory Press, 2001: 371-378.
- [23] Rychlik W. Oligo 7 primer analysis software. *Methods in Molecular Biology*, 2007, 402: 35-60.
- [24] Thompson J D, Gibson T J, Plewniak F, Jeanmougin F, Higgins D G. The Clustal _X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 1997, 25: 4876-4882.
- [25] Mathews D H. RNA secondary structure analysis using RNA structure. *Current Protocols in Bioinformatics*, 2006, 12 (6):1-14.
- [26] Kumar S, Tamura K, Nei M. MEGA: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, 2004, 5: 150-163.

- [27] Swofford D L. PAUP*: Phylogenetic analysis using parsimony (and other methods). Sinauer Associates Incorporated, 2003, 1-9.
- [28] Nikaido M. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *Journal of Molecular Evolution*, 2001, 53: 508-516.
- [29] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 1985, 39: 783-791.
- [30] Hedges S B. The number of replications needed for accurate estimation of the bootstrap *p* value in phylogenetic studies. *Molecular Biology and Evolution*, 1992, 9: 366-369.
- [31] Huelsenbeck J P, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 2001, 17(8): 754-755.
- [32] Posada D, Crandall K A. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 1998, 14: 817-818.
- [33] Posada D, Buckley T R. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 2004, 53: 793-808.
- [34] Rannala B & Yang Z. Probability distribution of molecular evolution trees: a new method of phylogenetic inference. *Journal of Molecular Evolution*, 1996, 43: 304-311.
- [35] Leache A D, Reeder T W. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Systematic Biology*, 2002, 51: 44-68.
- [36] Delarbre C, Rasmussen A S, Arnason U, Gachelin G. The complete mitochondrial genome of the hagfish *Myxine gluriosa*: unique features of the control region. *Journal of Molecular Evolution*, 2001, 53: 634-641.
- [37] Xu X, Arnason U. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene*, 1994, 148 (2): 357-362.
- [38] Sumida M, Kanamori Y, Kaneda H, Kato Y, Nishioka M, Hasegawa M. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*. *Genes and Genetic Systems*, 2001, 76:311-325.
- [39] Sano N, Kurabayashi A, Fujii T, Hiromichi Yonekawa, Masayuki Sumida. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the bellring frog, *Buergeria buergeri* (family Rhacophoridae). *Genes and Genetic Systems*, 2004, 79(3):151-163.
- [40] Wu X B, Wang Y Q, Zhou K Y, ZHU W G, NIE J S, WANG C L. Complete mitochondrial DNA sequence of Chinese alligator, *Alligator sinensis*, and phylogeny of crocodiles. *Chinese Science Bulletin*, 2003, 48(19): 2050-2054.
- [41] Janke A, Arnason U. The complete mitochondrial genome of Alligator mississippiensis and the separation between recent archosauria (birds and crocodiles). *Molecular Biology and Evolution*, 1997, 14: 1266-1272.
- [42] Janke A, Erpenbeck D, Nilsson M, Arnason U. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. *The Royal Society Proceedings B*, 2001, 268: 623-631.
- [43] Sun Y, Ma F, Xiao B, Zheng J J, Yuan X D, Tang M G, Wang L, YU Y F, Li Q W. The complete mitochondrial genomes sequences of *Asio flammeus* and *Asio otus* and comparative analysis. *Science in China Series C*, 2004, 47(6): 510-520.
- [44] Hu M, Chilton N B, Gasser R B. The mitochondrial genome of *Strongyloides stercoralis* (Nematoda) -idiosyncratic gene order and evolutionary implications. *International Journal for Parasitology*, 2003, 33:1393-1408.
- [45] Ursing B M, Arnason U. The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *Journal of Molecular Evolution*, 1998, 47:302-306.
- [46] Gissi C, Gullberg A, Arnason U. The complete mitochondrial DNA sequence of the rabbit (*Oryctolagus cuniculus*). *Genomics*, 1998, 50: 161-169.
- [47] Wei L, Wu X B, Jiang Z G. The complete mitochondrial genome structure of snow leopard *Panthera uncial*. *Molecular Biology Reports*, 2009, 36: 871-878.
- [48] Masuda R, Lopez J V, Pecon S J, Yuhki N, O'Brien S J. Molecular phylogeny of mitochondrial cytochrome *b* and 12S rRNA sequences in the Felidae: ocelot and domestic cat lineages. *Molecular Phylogenetics and Evolution*, 1996, 3:351-365.
- [49] Vila C, Amorim J R, Leonard J A, Posada D, Castroviejo J, Petrucci-Fonseca F, Crandall K A, Ellegren H, Wayne R K. Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular Ecology*, 1999, 8:2089-2103.
- [50] Aggarwal R K, Ramadevi J, Singh L. Ancient origin and evolution of the Indian wolf: evidence from mitochondrial DNA typing of wolves from Trans-Himalayan region and Pennisular India. *Genome Biology*, 2003, 4(6): 6-16.

参考文献:

- [20] 罗里杨, 金煜, 杨公社. 中国狼的种内系统发育关系与亚种分化. *东北林业大学学报*, 2003, 31(5): 81-83.
- [21] 孟超, 张洪海, 陈玉才. 中国狼线粒体全基因组序列分析. *中国生物化学与分子生物学报*, 2008, 24(12): 1170-1176.