

棘球属绦虫线粒体基因组全序列生物信息学分析

张耀刚^{1,2}, 马艳艳¹, 曹得萍², 姜博璠³, 樊海宁^{1,2}

摘要:目的 利用生物信息学技术分析棘球属绦虫线粒体基因组全序列碱基组成、基因的结构组成及排列、蛋白基因核苷酸序列变异位点、密码子使用情况及偏好性、系统发育等,为研究棘球属绦虫系统起源、演化、分类及亲缘关系等提供理论基础。方法 从 GenBank 数据库下载 12 种棘球属绦虫线 mtDNA 全序列,并以猪带绦虫作为外类群,构建系统发育树。利用 OGDRAW 在线分析网站及 Vector NTI Express 软件,分析 mtDNA 的组成及排列情况。用 Clust-X 软件对多个相似序列进行多重序列比对分析;使用 Mega4.0 软件选择邻接法构建进化树,并分析密码子使用情况;利用 RNAfold 在线预测网站使用 minimum free energy (MFE) and partition function 算法预测 l-rRNA 和 s-rRNA 二级结构。结果 除 Eg G1 外,棘球属绦虫属 mtDNA 有 36 个编码基因,包括 22 个转运 RNA 基因、12 个蛋白基因、2 个核糖体 RNA 基因;但 Eg G1 mtDNA 最小,只有 30 个编码基因,在起始编码区缺少 6 个编码转运 RNA 的基因,且第一个编码基因不是 ND5,而是 COX3;编码蛋白的基因核苷酸序列变异率为 27.9%~42.7%,其中 COX1 最为保守,ND5 变异率最大达到 42.7%;棘球属绦虫线粒体蛋白质编码基因起始密码子除 atg,也有一些蛋白质以 gtg 作为起始密码子,终止密码子以 taa 和 tag 常见,但也有以 ttt 作为终止密码子的;棘球属绦虫使用的密码子为密码表 9,使用频率最高的密码子是 UUG(2.72%),频率最低的是 CUC(1%)、CGC(1%),编码亮氨酸、精氨酸的密码子最多达 6 个,编码甲硫氨酸、色氨酸最少只有一个,亮氨酸也是棘球属绦虫最偏好的氨基酸达到 6%;棘球属绦虫核糖体基因有两个长度大小分别为 977~985 bp、700~727 bp,两个基因的位置十分靠近中间只隔一个 trnC 基因;系统进化树中 Ev、Eo 单独为一枝,Em、Es 及 Eg G1、Ef 形成姐妹枝,细粒棘球绦虫 G4、G5、G6、G7、G8、G10 亚型聚为一枝,进化距离较近。结论 本研究将为棘球属绦虫 mtDNA 的研究、分子进化和分子诊断等方面提供诸多信息,并为种系鉴定起到一定的指导作用。

关键词:棘球属绦虫;线粒体基因组;生物信息学分析;系统发育

中图分类号:R383.3

文献标识码:B

文章编号:1002-2694(2019)03-0271-07

Bioinformatic analyses on sequences of the complete mitochondrial genomes of *Echinococcus* genus

ZHANG Yao-gang^{1,2}, MA Yan-yan¹, CAO De-ping², JIANG Bo-fan³, FAN Hai-ning^{1,2}

(1. Central Laboratory of Qinghai University Affiliated Hospital, Xining 810000, China;

2. Qinghai Province Research Key Laboratory for echinococcosis, Xining 810000, China;

3. Qinghai University, Xining 810000, China)

Abstract: By means of bioinformatics, the whole sequence base composition, structure and arrangement of genes, nucleotide sequence variation sites of protein genes, codon usage and preference, phylogeny, etc., were analyzed in *Echinococcus*. It will provide a theoretical basis for studying the origin, evolution, classification and phylogenetic relationship of *Echinococcus* genus. The whole mtDNA sequence in 12 *Echinococcus* genus was downloaded from GenBank database, and the phylogenetic

tree was constructed with *Taenia solium* as an external group. The composition and arrangement of mtDNA were analyzed by OGDRAW online analysis website and Vector NTI Express software. Clust-X software was used to analyze multiple sequence alignment of similar sequences; Mega4.0 software selection adjacency method was used to construct evolutionary tree, and the use of codon was analyzed at last. The secondary structure of l-rRNA and s-rRNA was predicted by minimum free energy (MFE) and partition function algorithm using

青海大学附属医院中青年项目(No. ASRF-2018-YB-01)、2017 国家重点研发计划“精准医学研究”专项(No. 2017YFC0909900)、青海省应用基础研究项目(No. 2014-ZJ-719)

通讯作者:樊海宁, Email: fanhaining@medmail.com.cn;

ORCID: 0000-0001-7791-5725

作者单位: 1. 青海大学附属医院中心实验室, 西宁 810001;

2. 青海省包虫病研究重点实验室, 西宁 810001;

3. 青海大学, 西宁 810016

RNAfold online prediction website. Results showed that the size of mtDNA of *Echinococcus* genus was 13.5 kB–13.8 kB. The genome was rich in A and T, in which A content was up to about 48%, followed by T 20%, and the coding region accounted for about 73% of the whole genome sequence. Apart from Eg G1, the mtDNA of *Echinococcus* genus contains 36 coding genes which include 22 transporter RNA genes, 12 protein genes and 2 ribosomal RNA genes. However, the genome of E.g. G1 mtDNA was the smallest with only 30 coding genes, lacking six tRNA genes in the initial coding region, and the first coding gene was COX3 instead of ND5. The nucleotide sequence variation rate of the gene encoding protein ranged from 27.9% to 42.7%, of which COX1 was the most conservative and ND5 had the highest variation rate of 42.7%. In addition to “atg”, some proteins use “gtg” as the starting codon, while “taa” and “tag” were common for termination codons, but “ttt” was also used for termination codon. Codon Table 9 was used by *Echinococcus* genus, of which UUG (2.72%) was the most frequently used codon, while CUC (1%) and CGC (1%) were the lowest. There were six codons encoding leucine and arginine, only one codon encoding methionine and tryptophan. Leucine was also the most preferred amino acid of *Echinococcus* genus with 6%. The ribosomal genes of *Echinococcus* genus had two lengths ranging from 977-985 bp and 700-727 bp respectively, and the positions of the two genes were very close, separated by one trnC gene; in phylogenetic tree, Ev and Eo were single branch while Em, E.s and Eg G1, Ef form sister branches, and The G4, G5, G6, G7, G8, and G10 subtypes of *Echinococcus granulosus* were clustered into one branch with a relatively close evolutionary distance. In conclusion, this study will provide a lot of information for the research, molecular evolution and molecular diagnosis of mtDNA in *Echinococcus* genus, which plays a guiding role in phylogenetic identification, drug and vaccine development in the future.

Keywords: *Echinococcus* genus; mitochondrial genome; bioinformatics analysis; phylogeny

Supported by the Youth and Middle Age Project of Qinghai University Affiliated Hospital (No.ASRF-2018-YB-01), National Key R&D Program "Precision Medical Research" in 2017 (No.2017YFC 0909900) and Basic Applied Research Project of Qinghai Province (No.2014-ZJ-719)

Corresponding author: Fan Hai-Ning, Email: fanhaining@medmail.com.cn

近年来学者对于棘球绦虫的分类做了大量研究,由于棘球绦虫具有独特的繁殖方式:成虫雌雄同体,以自体受精为主;中绦期为无性繁殖。这种双向发育和繁殖模式造就了棘球绦虫的生物多样性,加速了遗传一致性种群的形成^[1],目前至少有 10 种亚种的报道,学者们通过分析细粒棘球绦虫 mtDNA 全基因组序列,将细粒棘球绦虫 G1-G3 基因型归为狭义细粒棘球绦虫(*Echinococcus granulosus sensu stricto*)、G4 基因型归为马棘球绦虫(*Echinococcus equinus*)、G5 基因型归为奥式棘球绦虫(*Echinococcus ortleppi*)、G6-G10 基因型归为加拿大棘球绦虫(*Echinococcus canadensis*)及细粒棘球绦虫狮子株归为狮棘球绦虫(*Echinococcus felidis*)^[2]。但不少亚种能否单独存在尚存在争议。线粒体基因是典型的母性遗传,因其具有结构简单稳定、无重组、易分离纯化、由于受到线粒体内氧化环境的影响而容易发生突变,因此其进化速率快,是研究寄生虫分子进化的理想工具。在构建亲缘关系相近物种的进化树时比使用核 DNA 序列更有优势。利用不同虫株(基因型)在幼虫形态、致病性、宿主范围、流行病学意义等方面存在的差异,通过分析线粒体基因的种类和数量、遗传密码子进化、同义密码子使用、RNA 转录成熟机制及碱基组成的偏向性、基因重叠及重

排现象等问题^[3],能反映寄生虫的不同进化路线,因而被广泛用于生物间系统起源、演化、分类及亲缘关系的研究^[4]。同时对疫苗、诊断试剂及抗虫药物的研制和开发具有重要意义^[5-6]。本研究通过分析棘球属绦虫线粒体基因组序列的基因结构、碱基组成、遗传密码子及分子系统发育,以期为种系鉴定提供理论参考。

1 材料与方法

1.1 序列获取 从 NCBI GenBank 数据库下载棘球属绦虫线粒体基因组 DNA 全序列,包括多房棘球绦虫(*Echinococcus multilocularis*, Em)、福氏棘球绦虫(*Echinococcus vogeli*, Ev)、少节棘球绦虫(*Echinococcus oligarthrus*, Eo)、狮棘球绦虫(*Echinococcus felidis*, Ef)、石渠棘球绦虫(*Echinococcus shiquicus*, Es)及细粒棘球绦虫(*Echinococcus granulosus*, Eg)多个虫株(G1、*Echinococcus equinus* G4、*Echinococcus ortleppi* G5、*Echinococcus canadensis* G6、*Echinococcus canadensis* G7、*Echinococcus canadensis* G8、*Echinococcus canadensis* G10),并以猪带绦虫(*Taenia solium*, Ts)作为外类群。

1.2 mtDNA 的组成及排列使用 OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw>).

html)在线分析网站及 Vector NTI Express 软件。

1.3 序列分析及分子进化树的构建 用 Clust-X 软件对多个相似序列进行多重序列比对分析。使用 Mega4.0 软件选择邻接法(Neighbor-joining method,NJ)构建进化树,并做 Bootstrap 检验,重复次数设置为 10000,选择 Bootstrap 检验。密码子使用情况使用 Mega4.0 软件分析。

1.4 核糖体基因分析及二级结构预测 l-rRNA 和 s-rRNA 二级结构预测使用 RNAfold 在线预测网站(http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi),选择 minimum free energy (MFE) and partition function 算法。

2 结果

2.1 线粒体基因组基因的组成及排列 除 Eg G1

外,棘球绦虫属线粒体基因组有 36 个编码基因,包括 22 个转运 RNA 基因、12 个蛋白基因、2 个核糖体 RNA 基因。其中 Eo、Es、Ev、Eg G1、Eg G4、Ef 线粒体 ND4L-ND4 基因基因序列存在 2-30 bp 序列的重叠。基因排列次序为:trnY-trnL-trnS-trnL-trnR-ND5-trnG-COX3-trnH-CYTB-ND4L-ND4-trnQ-trnF-trnM-ATP6-ND2-trnV-trnA-trnD-ND1-trnN-trnP-trnI-trnK-ND3-trnS-trnW-COX1-trnT-rrnL-trnC-rrnS-COX2-trnE-ND6。但 Eg G1 线粒体基因组最小,只有 30 个编码基因,在起始编码区缺少 6 个编码转运 RNA (trnY-trnL-trnS-trnL-trnR-trnG)的基因。且第一个编码基因不是 ND5,而是 COX3,ND5 变为最后一个。*Em*、*EgG1* mtDNA 全序列见图 1、图 2。

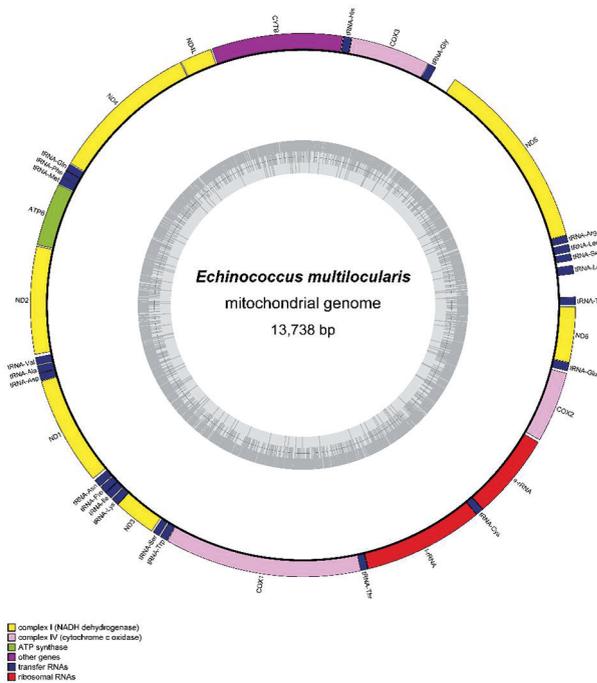


图 1 多房棘球绦虫线粒体基因组全基因组序列

Fig.1 *Echinococcus multilocularis* mitochondrion, complete genome

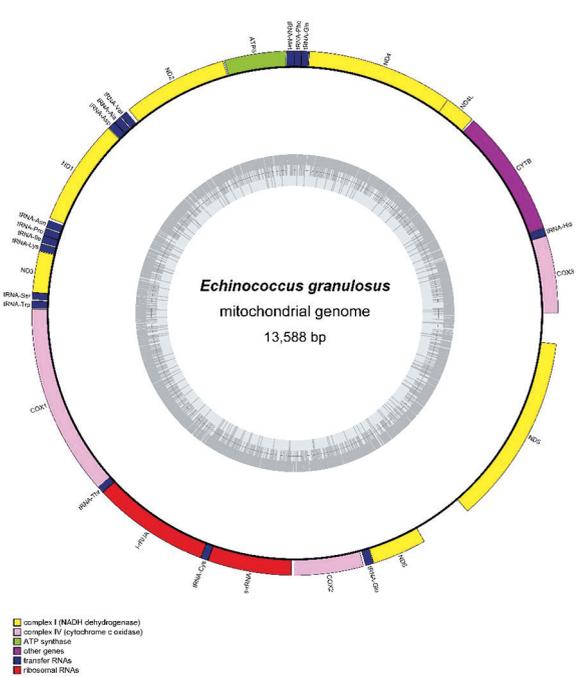


图 2 细粒棘球绦虫 G1 型线粒体基因组全基因组序列

Fig.2 *Echinococcus granulosus* G1 mitochondrion, complete genome

2.2 棘球属绦虫线粒体基因组的蛋白基因核苷酸

序列变异位点分析 12 种棘球属绦虫线粒体基因组的蛋白基因核苷酸序列变异位点分析结果显示,蛋白基因核苷酸序列变异率为 27.9%~42.7%,其中 COX1 最为保守,这也是在棘球属种系鉴定中常选择 COX1 基因作为分子标记物的原因之一,其次

为 ND3、COX2 变异率分别为 27.9%、29.7%,另外 ND4、ATP6、ND6、ND5 基因进化速率较快,其中以 ND5 变异率最大达到 42.7%,因此,ND4、ATP6、ND6、ND5 基因也可以作为 COX1 基因的辅助分子标记(表 1)。

表 1 棘球属绦虫线粒体基因组的蛋白编码基因核苷酸序列变异位点分析

Tab.1 Nucleotide sequence variation of protein-coding genes in mitochondrial genome of *Echinococcus* genus

基因名称	ND5	COX3	CYTB	ND4L	ND4	ATP6	ND2	ND1	ND3	COX1	COX2	ND6
总位点数	1578	648	1068	276	1264	516	885	897	348	1608	582	456
不变位点	904	443	759	183	777	315	573	589	251	1232	409	273
变异位点	674	205	309	84	483	198	309	308	97	376	173	183
单态突变点	264	72	123	26	176	72	107	103	35	138	66	60
简约性信息位点	410	133	186	58	307	126	202	205	62	238	107	123
变异位点百分比(%)	42.7	31.6	28.9	33.7	38.5	39.0	35.3	34.3	27.9	23.4	29.7	40.1

2.3 转录密码子使用情况 与大多数真核生物一样,棘球属线虫线粒体蛋白基因起始密码子主要为 *atg*,但也有一些蛋白质以 *gtg* 作为起始密码子,COX2 在 12 种棘球属绦虫中均以 *gtg* 作为起始密

码子,有些蛋白质只在部分虫株中以 *gtg* 作为起始密码,例如 ND4L、ND4、ND2、ND1。终止密码子以 *taa* 和 *tag* 常见,但也有以 *tta* 作为终止密码子(表 2)。

表 2 棘球属绦虫线粒体编码蛋白基因起始密码子和终止密码子

Tab.2 Initiation codon and termination codon of mitochondrial protein gene in *Echinococcus* genus

种名	线粒体基因组蛋白编码基因											
	ND5	COX3	CYTB	ND4L	ND4	ATP6	ND2	ND1	ND3	COX1	COX2	ND6
Em	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/taa</i>
Eo	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Es	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>gtg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>gtg/taa</i>	<i>atg/tag</i>
Ev	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tta</i>	<i>gtg/tag</i>	<i>atg/taa</i>
Eg G1	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>gtg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Eg G4	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Eg G5	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Eg G6	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Eg G7	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Eg G8	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Eg G10	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/tag</i>	<i>gtg/tag</i>	<i>atg/tag</i>
Ef	<i>atg/taa</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>gtg/taa</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/tag</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>atg/taa</i>	<i>gtg/tag</i>	<i>atg/tag</i>

2.4 蛋白翻译密码子使用情况 棘球属绦虫属使用的密码子为棘皮类和扁形虫线粒体密码(transl_table=9),与标准密码子表的区别为:在标准密码子表中 AAA 编码 Lys,AGA、AGG 编码 Arg,UGA 为终止密码子,而棘球属绦虫中 AAA 编码 Asn,AGA、AGG 编码 Ser,UGA 编码 Trp。在棘球属绦虫中编码氨基酸的密码子使用频率最高的是 UUG

(2.72%),频率最低的是 CUC(1%)、CGC(1%),编码亮氨酸 L(UUA、UUG、CUU、CUC、CUA、CUG)、精氨酸 R(CGU、CGC、CGA、CGG、AGA、AGG)的密码子最长达 6 个,编码甲硫氨酸 M(AUG)、色氨酸 W(UGG)最少只有 1 个。亮氨酸 L 也是棘球属绦虫最偏好的氨基酸达到 6%(表 3)。

表 3 蛋白编码密码子使用情况
Tab.3 Protein coding codon usage

密码子	频数/%	密码子	频数/%	密码子	频数/%	密码子	频数/%
UUU(F)	484.8(1.77)	UCU(S)	78.7(1.62)	UAU(Y)	228.7(1.70)	UGU(C)	219.2(1.65)
UUC(F)	63.4(0.23)	UCC(S)	17.2(0.35)	UAC(Y)	40.2(0.30)	UGC(C)	46.3(0.35)
UUA(L)	225.9(2.04)	UCA(S)	36.2(0.74)	UAA(*)	77.7(0.99)	UGA(*)	86.7(1.10)
UUG(L)	301.7(2.72)	UCG(S)	29.0(0.60)	UAG(*)	71.3(0.91)	UGG(W)	132.2(1.00)
CUU(L)	53.2(0.48)	CCU(P)	23.0(1.81)	CAU(H)	40.6(1.62)	CGU(R)	37.3(1.35)
CUC(L)	10.9(0.10)	CCC(P)	4.2(0.33)	CAC(H)	9.6(0.38)	CGC(R)	2.8(0.10)
CUA(L)	30.9(0.28)	CCA(P)	13.1(1.03)	CAA(Q)	12.1(0.79)	CGA(R)	8.9(0.32)
CUG(L)	42.4(0.38)	CCG(P)	10.7(0.84)	CAG(Q)	18.3(1.21)	CGG(R)	18.1(0.66)
AUU(D)	196.8(1.85)	ACU(T)	44.8(2.06)	AAU(N)	93.2(1.75)	AGU(S)	113.3(2.33)
AUC(D)	25.0(0.24)	ACC(T)	11.0(0.50)	AAC(N)	13.6(0.25)	AGC(S)	17.8(0.37)
AUA(D)	96.6(0.91)	ACA(T)	18.4(0.85)	AAA(K)	36.1(0.81)	AGA(R)	39.3(1.42)
AUG(M)	107.3(1.00)	ACG(T)	12.9(0.59)	AAG(K)	53.0(1.19)	AGG(R)	59.2(2.15)
GUU(V)	297.7(2.15)	GCU(A)	45.4(2.01)	GAU(D)	82.8(1.66)	GGU(G)	158.4(2.36)
GUC(V)	35.0(0.25)	GCC(A)	14.8(0.65)	GAC(D)	17.1(0.34)	GGC(G)	21.4(0.32)
GUA(V)	77.9(0.56)	GCA(A)	14.0(0.62)	GAA(E)	38.5(0.87)	GGA(G)	34.0(0.51)
GUG(V)	143.8(1.04)	GCG(A)	16.1(0.71)	GAG(E)	50.5(1.13)	GGG(G)	54.4(0.81)

* 终止密码子

2.5 核糖体(rRNA)基因的二级结构 棘球属绦虫核糖体基因有两个分别为 l-rRNA 及 s-rRNA,长度大小分别为 977~985 bp、700~727 bp,两个基因的位置十分靠近中间只隔一个 *trnC* 基因。通过

RNAfold 在线预测网站预测棘球绦虫 rRNA 结构类似,形成较多的茎环结构。多房棘球绦虫 rRNA 结构见图 3。



图 3 多房棘球绦虫核糖体(rRNA)基因的二级结构

Fig.3 Secondary structure of the ribosome (rRNA) gene of *Echinococcus multilocularis*

2.6 线粒体基因组序列系统进化分析 基于 12 种棘球绦虫线粒体基因组全序列,以 *T.s* 作为外类群,使用邻接法构建系统进化树。由图 5 可以看出 *Ev*、*Eo* 单独为一枝,*Em*、*Es* 及 *Eg* G1、*Ef* 形成姐妹枝。细粒棘球绦虫 G4、G5、G6、G7、G8、G10 亚型聚为一枝,进化距离较近(见图 4)。

3 讨论

目前棘球属绦虫种及亚种的分类存在着争议,自从多房棘球绦虫、少节棘球绦虫、福氏棘球绦虫立为独立种之后,争论渐趋平息,目前比较公认的棘球属绦虫共有 4 种,但近年,我国学者肖宁等先后从高原鼠兔(*Ochotona curzoniae*)和藏狐(*Vulpes ferri-*

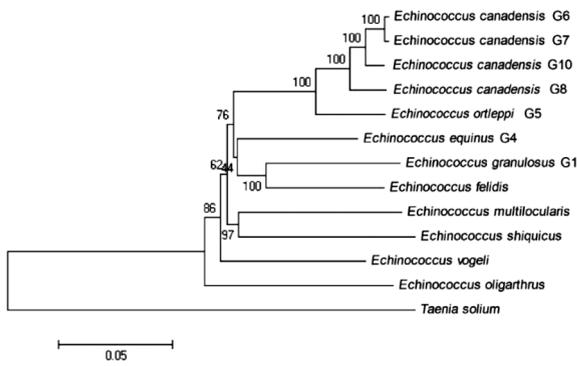


图4 棘球属绦虫系统发育树

Fig.4 Phylogenetic tree of *Echinococcus* genus

lata)分离出一种棘球绦虫,其成虫和幼虫形态、分子遗传特点、寄生宿主范围和地理分布、物种进化等方面与其他棘球属绦虫存在较大差异,具备种间差异条件,并以其首次发现地——四川石渠县命名为石渠棘球绦虫^[7-8]。但其能否作为新的种还存在许多疑点,石渠棘球绦虫是否为细粒棘球绦虫或多房棘球绦虫在非适宜宿主中适应性的改变不得而知。由于棘球属绦虫特别是细粒棘球绦虫种内变异现象非常普遍^[9],根据线粒体 COX1 和 NAD1 基因部分序列的差异,细粒棘球绦虫分为 G1-G10 10 种不同基因型^[1,10-12]。

随着新的寄生宿主^[13]及生物学特征^[14]被发现,这样的分类依然不尽完善,因此本研究通过分析 12 种报道的棘球属绦虫的线粒体基因组序列,为深入了解棘球属绦虫,本文从棘球属绦虫线粒体基因组序列碱基组成、基因结构与排列、密码子使用及偏好、系统发育等进行分析。通过分析发现棘球属线粒体主要有三部分组成包括 22 个转运 RNA 基因、12 个蛋白基因、2 个核糖体 RNA 基因,但 E.g G1 线粒体基因组发生了重排,只有 30 个编码基因,且第一个编码基因不是 ND5,而是 COX3,ND5 变为最后一个,重排是种系进化中的稀有事件,重排的规律性可为重建物种演化历史提供重要信息。且 E.g G1 缺少 6 个编码转运 RNA,无法转运酪氨酸、亮氨酸、丝氨酸、精氨酸、甘氨酸,这些氨基酸可能有其他转运 RNA 转运或由宿主提供。

目前对于棘球属绦虫基因型分类常用的工具包括核基因和 mtDNA,核基因主要包括核糖体 DNA 第一、二内转录间隔区基因 ITS1、ITS2,肌动蛋白 III,延长因子 1A,转化生长因子,硫氧还蛋白过氧化物酶,钙网蛋白,疏水性类弹性蛋白,抗原 B 1, Eg9 和 Eg16 等,而线粒体基因主要有 *cox1*、*nad1*、*nad2*、*nad3*、*cytb*、*ATP6*、12s rRNA 和 16s rRNA

等^[15-16],一般认为线粒体基因序列进化较快,种系发育关系分析方面比核基因更有效,随着近年来学者们对于棘球属线粒体基因测序数据的不断积累,棘球绦虫 mtDNA 数据库较核基因更完善,在棘球属绦虫分类中更具有优势^[17]。本研究通过多序列比较也发现,12 个线粒体基因编码的基因中 COX I 最为保守,ND5 变异率最大达到 42.7%,另外 ND4、ATP6、ND6 基因进化速率较快,提示除 COX I 基因外,ND4、ATP6、ND6、ND5 也可辅助作为棘球绦虫鉴定的分子标记。

同时,棘球绦虫属寄生虫缺少高等生物拥有的 ATP8 基因,这可能是棘球属绦虫在寄生的环境中丢弃了部分“无用”的基因^[18],进化为更高级的生存方式,这一点在多房棘球绦虫生长特性中表现得更为明显,因为在中间宿主人中无法找到虫体,仅有一层叫生发层的细胞即可实现无限增殖,在寄生虫疾病中极为罕见^[19];起始密码中除以 *atg* 作为起始密码外,一些进化较快的基因例如 COX2 及 ND 部分亚基还以 *gtg* 作为起始密码,终止密码子中甚至有以 *ttt* 作为终止密码的。如此一个基因的终止密码中的碱基也可能是另一个基因的起始密码中的碱基,使得线粒体碱基的利用度达到最大,结构更为紧凑。棘球绦虫属使用频率最高的蛋白编码密码子是 UUG,频率最低的是 CUC 和 CGC,编码亮氨酸、精氨酸的密码子最多达 6 个,而编码甲硫氨酸(AUG)、色氨酸(UGG)最少只有 1 个。亮氨酸也是棘球属绦虫最偏好的氨基酸达到 6%,这也反映出棘球属绦虫对于氨基酸具有明显的偏好性。

使用邻接法构建系统进化树可以看出 E.v、E.o 单独为一枝,E.m、E.s 及 E.g G1、E.f 形成姐妹枝,细粒棘球绦虫 G4、G5、G6、G7、G8、G10 亚型聚为一枝,进化距离较近,这与 Wassermann M 等人的研究结果较为一致^[20]。在发现石渠棘球绦虫前,因其形态学上与 Em 相近似,则被误认为是多房棘球绦虫的变异种,由线粒体进化距离上看两者亲缘关系也较近。总之,利用线粒体基因研究棘球绦虫系统起源、演化、分类及亲缘关系,离不开对于线粒体基因的充分了解,通过本文分析期望能为进一步认识棘球属线粒体基因组提供一定的参考。

利益冲突:无

引用本文格式:张耀刚,马艳艳,曹得萍,等.棘球属绦虫线粒体基因组全序列生物信息学分析[J].中国人兽共患病学报,2019,35(3):271-277. DOI: 10.3969/j.issn.1002-2694.2019.00.013

参考文献:

- [1] Addy F, Wassermann M, Kagendo D, et al. Genetic differentiation of the G6/7 cluster of *Echinococcus canadensis* based on mitochondrial marker genes[J]. *Int J Parasitol*, 2017, 47(14): 923-931. DOI: 10.1016/j.ijpara.2017.06.003
- [2] Kinkar L, Laurimäe T, Sharbatkhori M, et al. New mitogenome and nuclear evidence on the phylogeny and taxonomy of the highly zoonotic tapeworm *Echinococcus granulosus* sensu stricto [J]. *Infect Genet Evol*, 2017, 52:52-58. DOI:10.1016/j.meegid.2017.04.023
- [3] 罗厚强. 林芝藏猪消化道寄生虫调查及蛔虫、细颈囊尾蚴线粒体基因组学研究[D]. 武汉:华中农业大学, 2017.
- [4] 盛军庆, 林巧惠, 王军花, 等. 池蝶蚌线粒体基因组全序列分析[J]. *水生生物学报*, 2014, (2): 320-327. DOI:10.7541/2014.46
- [5] 李立, 娄忠子, 闫鸿斌, 等. 棘球属绦虫分子种系发生研究进展[J]. *中国人兽共患病学报*, 2014, 31(11): 1155-1162. DOI: 10.3969/j.issn.1002-2694.2014.11.015
- [6] 郝桂英. 线粒体基因在带属绦虫分子分类中的研究进展[J]. *动物医学进展*, 2014, (1): 94-98. DOI: 10.16437/j.cnki.1007-5038.2014.01.019
- [7] Wang X, Liu J, Zuo Q, et al. *Echinococcus multilocularis* and *Echinococcus shiquicus* in a small mammal community on the eastern Tibetan Plateau: host species composition, molecular prevalence, and epidemiological implications[J]. *Parasit Vectors*, 2018, 11(1): 302. DOI: 10.1186/s13071-018-2873-x
- [8] Ma J, Wang H, Lin G, et al. Surveillance of *Echinococcus* isolates from Qinghai, China[J]. *Vet Parasitol*, 2015, 207(1/2): 44-48. DOI: 10.1016/j.vetpar.2014.11.012
- [9] Laurimäe T, Kinkar L, Moks E, et al. Molecular phylogeny based on six nuclear genes suggests that *Echinococcus granulosus* sensu lato genotypes G6/G7 and G8/G10 can be regarded as two distinct species[J]. *Parasitology*, 2018:1-9. DOI: 10.1017/S0031182018000719
- [10] Cucher MA, Macchiaroli N, Baldi G, et al. *Cystic echinococcosis* in South America: systematic review of species and genotypes of *Echinococcus granulosus* sensu lato in humans and natural domestic hosts[J]. *Trop Med Int Health*, 2016, 21(2): 166-175. DOI: 10.1111/tmi.12647
- [11] Khademvatan S, Majidiani H, Foroutan M, et al. *Echinococcus granulosus* genotypes in Iran: a systematic review[J]. *J Helminthol*, 2018: 1-8. DOI: 10.1017/S0022149X18000275
- [12] Alvarez RC, Romig T, Lightowlers M W. *Echinococcus granulosus* sensu lato genotypes infecting humans—review of current knowledge[J]. *Int J Parasitol*, 2014, 44(1): 9-18. DOI: 10.1016/j.ijpara.2013.08.008
- [13] Hodzic A, Alic A, Supic J, et al. *Echinococcus ortleppi*, the cattle strain in a crested porcupine (*Hystrix cristata*): A new host record[J]. *Vet Parasitol*, 2018, 256: 32-34. DOI: 10.1016/j.vetpar.2018.05.004
- [14] Lymbery AJ, Jenkins EJ, Schurer JM, et al. *Echinococcus canadensis*, *E. borealis*, and *E. intermedius*. What's in a name[J]. *Trends Parasitol*, 2015, 31(1): 23-29. DOI: 10.1016/j.pt.2014.11.003
- [15] Nikmanesh B, Mirhendi H, Mahmoudi S, et al. Multilocus sequence analysis of *Echinococcus granulosus* strains isolated from humans and animals in Iran[J]. *Exp Parasitol*, 2017, 183:50. DOI: 10.1016/j.exppara.2017.10.002
- [16] Héctor G Avila, Santos GB, Cucher MA, et al. Implementation of new tools in molecular epidemiology studies of *Echinococcus granulosus* sensu lato in South America[J]. *Int J Parasitol*, 2017, 66(3):250-257. DOI: 10.1016/j.parint.2017.02.001
- [17] Marinova I, Spiliotis M, Wang J, et al. Molecular characterization of *Echinococcus granulosus* isolates from Bulgarian human cystic echinococcosis patients[J]. *Parasitology Research*, 2017, 116(3):1043-1054. DOI: 10.1007/s00436-017-5386-1
- [18] Tsai IJ, Zarowiecki M, Holroyd N, et al. The genomes of four tapeworm species reveal adaptations to parasitism[J]. *Nature*, 2013, 496(7443): 57-63. DOI: 10.1038/nature12031
- [19] Koziol U, Rauschendorfer T, Zanon RL, et al. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*[J]. *Evodevo*, 2014, 5(1): 10. DOI: 10.1186/2041-9139-5-10
- [20] Wassermann M, Woldeyes D, Gerbi BM, et al. A novel zoonotic genotype related to *Echinococcus granulosus* sensu stricto from southern Ethiopia[J]. *Int J Parasitol*, 2016, 46(10): 663-668. DOI:10.1016/j.ijpara.2016.04.005

收稿日期:2018-09-05 编辑:王晓欢