



奇异双身虫线粒体基因组测定及其密码子偏好性分析

石彩霞¹, 王瑾璞¹, 张丽¹, 刘彦君¹,
魏念文¹, 岳城¹, 郝翠兰^{1,2*}

(1. 新疆农业大学动物医学学院, 新疆乌鲁木齐 830052;

2. 新疆草食动物新药研究与创制重点实验室, 新疆乌鲁木齐 830052)

摘要: 为确定奇异双身虫线粒体基因组基本结构特征并探究奇异双身虫分类地位, 同时了解双身虫科物种的密码子偏好性, 实验采用二代高通量测序技术获得奇异双身虫线粒体基因组序列, 通过线粒体在线注释网站 MITOS 对线粒体基因组序列进行注释并通过 Phylo-Suite (version 1.2.2) 软件对其结构进行分析, 同时对奇异双身虫及其他 8 种双身虫科物种的线粒体蛋白质编码基因的密码子组成情况及密码子偏好性进行分析。结果显示, 奇异双身虫的线粒体基因组序列长度为 15 713 bp (大非编码区未测全), A+T 含量为 68.7%, 表现出明显的 AT 偏向。系统发育树结果显示, 基于现有数据双身虫科物种聚为一支, 支持双身虫科物种的单系性。对双身虫科物种的密码子使用偏好进行分析, 结果显示双身虫科有效密码子数 (ENC) 为 30.632~37.495, 相对同义密码子使用度 (RSCU) > 1 的密码子共有 18 个, 偏好以碱基 U(T) 结尾。PR2-plot 分析结果显示, T 的使用频率高于 A, G 的使用频率高于 C, 突变和选择等因素均可对密码子的使用偏好产生一定的影响。中性绘图分析显示, 奇异双身虫和马口鱼拟双身虫受突变压力的影响分别为 61.89% 和 51.31%, 其余 7 种双身虫受突变压力的影响均小于 50%。双身虫科密码子偏好性存在略微差异, 但都受突变压力、自然选择和碱基组成的共同影响。本研究可以为后续的系统进化研究提供重要依据。

关键词: 奇异双身虫; 线粒体基因组; 密码子偏好性; 系统发育

中图分类号: Q 959.152; S 941.5

文献标志码: A

双身虫是由两个未成熟的幼虫在鱼类鳃部相互接触, 然后逐渐结合到一起, 最后形成呈“X”形的性成熟成虫。双身虫隶属于扁形动物门 (Platyhelminthes) 单殖吸虫纲 (Monogenea) 铗钩虫目 (Mazocraeidea) 双身虫科 (Diplozoidae), 是一类重要的单殖吸虫物种。奇异双身虫 (*Diplozoon paradoxum*) 隶属于双身虫科 (Diplozoidae) 双身虫属 (*Diplozoon*), 是一种重要的单殖吸虫物种^[1]。主要分布于欧洲的波罗的海、黑海、里海等水域, 且

在欧鳊 (*Abramis brama*)、拟鲤 (*Rutilus rutilus*)、真鳊 (*Phoxinus phoxinus*) 等鱼类上均有发现。在中国, 仅在额尔齐斯河 (中国段) 东方欧鳊 (*A. brama orientalis*) 的鳃部有所发现。双身虫的种类很多, 包括云南拟双身虫 (*Paradiplozoon yunnanensis*)、马口鱼拟双身虫 (*P. opsariichthydis*)、鳌拟双身虫 (*P. hemiculteri*)、叶尔羌拟双身虫 (*P. yarkandense*)、西伯利亚拟双身虫 (*P. homoion*)、日本真双身虫 (*Eudiplozoon nipponicum*)、真双身虫未定种

收稿日期: 2023-11-23 修回日期: 2024-01-16

资助项目: 新疆维吾尔自治区重点研发任务专项 (2023B02012); 国家自然科学基金 (31960737); 新疆维吾尔自治区自然科学基金 (2021D01B56)

第一作者: 石彩霞, 从事水生动物病害防治研究, E-mail: 1051718934@qq.com

通信作者: 郝翠兰 (照片), 从事水生动物病害防治研究, E-mail: haocuilan@126.com



(*Eudiplozoon* sp. DZ-2018) 和华双身虫未定种 (*Sindiplozoon* sp. DZ-2018) 等。双身虫科目前公开的线粒体基因组包括云南拟双身虫、马口鱼拟双身虫、鳖拟双身虫等 8 种, 其中个别物种在非编码区未测全, 且双身虫属没有数据公开。双身虫线粒体基因组与其他扁形动物的线粒体基因组一样, 含有 36 个基因, 其中蛋白质编码基因有 12 个。

蛋白质编码基因的密码子包括 64 个三联密码子, 这些三联密码子有的属于同义密码子, 即编码相同氨基酸的密码子, 64 个三联密码子一共可以编码 20 种氨基酸^[2]。在密码子使用过程中, 会出现某些密码子使用频率高于其他同义密码子使用频率的现象, 称为密码子偏好性^[3]。密码子偏好性形成是多因素综合作用的结果, 早期针对密码子偏好性的解释, 主要是“选择-突变-漂移”和“中性理论”假说^[4]。随着对密码子偏好性的深入研究, 发现影响偏好性形成的因素很多^[5-7], 包括突变压力^[8]、自然选择^[8]、基因长度^[9]、编码区 GC 含量^[10]、编码蛋白质的结构与功能^[11]、内含子长度^[12]、环境压力^[13] 和种群大小^[14] 等。在这些因素中, 普遍认为突变压力和选择压力是密码子偏好性形成的关键原因^[15-16], 而两者中何者又占据主导是当前比较关注的点。

先前学者对奇异双身虫的研究主要集中在形态学、种群动态学等方面^[1, 17]。目前, 奇异双身虫线粒体基因组及双身虫密码子偏好性的比较工作尚未开展。本研究首次对奇异双身虫线粒体基因组进行测序和注释, 并对其核苷酸组成进行分析, 对分析该物种的密码子使用偏好。本研究补充了双身虫属线粒体基因组数据, 为双身虫科系统发育研究提供基础数据, 而且通过分析双身虫科物种密码子组成的各项指标, 分析研究密码子使用偏好及其影响因素, 以期对双身虫的分子进化和物种进化研究提供理论基础。

1 材料与方法

1.1 样品采集

奇异双身虫样本于 2021 年 7 月采集自额尔齐斯河流域 (185 团河段)(47°59'50"N, 85°41'40"E)。样品保存于 95% 的乙醇中, 经形态鉴定后保存于 -20 °C 冰箱。参照 DNA 提取试剂盒 (EasyPure® Genomic DNA Kit) 说明提取其总 DNA, 经 1% 琼

脂糖凝胶电泳检测 DNA 质量后送往南京派森诺基因科技有限公司进行高通量测序。

1.2 线粒体基因组的测序和组装

采用全基因组鸟枪法 (whole genome shotgun, WGS) 构建文库, 基于 Illumina Novaseq 平台对奇异双身虫 DNA 文库进行双末端 (Paired-end, PE) 测序, 插入大小为 400 bp 的目标片段。使用 FastQC 软件 (<http://www.bioinformatics.babaaham.ac.uk/projects/fastqc>) 检测原始数据的质量, 使用适配器去除数据过滤进行数据组装。通过采用 AdapterRemoval (version2) 去除 3'端的接头污染; 采用 SOAPec (version2.01) 软件基于 K-mer 分布对所有读序 (reads) 进行质量校正来获得高质量数据。采用 A5-miseq (version20150522) 和 SPAdes (version 3.9.0) 对高质量的二代测序数据进行从头拼装^[18-19], 构建 contig 和 scaffold 序列。使用 pilon (version 1.18)^[20] 软件对结果进行校正以得到最终的线粒体序列。

1.3 基因组注释

使用 MITOS Web Server^[21] (<http://mitos.bioinf.uni-leipzig.de/>) 对全序列进行初步分析, 遗传密码选择设置为棘皮动物和扁形动物, 其余设置按照 MITOS 设置的默认参数, 再通过近缘种比对来确定各基因的起始和终止位点。

1.4 系统发育分析

本研究选取 NCBI 上公布的 15 个代表种的线粒体基因组和本研究新测序的奇异双身虫线粒体基因组, 选择 *Obrimoposthia wandeli* 作为外类群 (表 1)。利用 PhyloSuite (version1.2.2) 提取单殖吸虫 36 个基因的核苷酸序列, 使用 MAFFT (version7.313) 对蛋白质编码基因 (PCGs) 的核苷酸序列进行多重序列比对^[22]。将处理好的数据串联成 12 个蛋白质编码基因序列的联合数据集, 利用 PhyloSuite 程序中的 PartitionFinder 2 来确定最优分区策略及其模型^[23], 使用贝叶斯法 (Bayesian inference, BI) 和最大似然法 (Maximum likelihood, ML) 构建系统发育树。BI 分析用贝叶斯推断方法构建系统发育树, 模型选择后设置 4 条马尔科夫 (MCMC) 链独立运行 2 000 000 代, 每隔 1 000 代进行 1 次采样, 前 25% 被舍弃, 使用贝叶斯后验概率 (BPP) 来检测节点的支持率。ML 分析采用 IQ-TREE^[24] 和 1 000 个超快引导进行。美化系统

表 1 本研究中分析的物种列表及其 GenBank 登录号

Tab. 1 The list of species analyzed in this study and their GenBank accession numbers

物种 species	长度/bp length	登录号 accession no.
大睾伪鲟缺虫 <i>Pseudochauhanea macrorchis</i>	15 031	NC_016950.1
<i>Neoheterobothrium hirame</i>	14 588	MN984338.1
真双身虫未定种 <i>Eudiplozoon</i> sp. DZ-2018	14 334	MG458328.1
日本真双身虫 <i>E. nipponicum</i>	17 328	NC_061193.1
云南拟双身虫 <i>P. yunnanensis</i>	14 829	NC_062047.1
马口鱼拟双身虫 <i>P. opsariichthydis</i>	15 385	MG458327.1
螯拟双身虫 <i>P. hemiculteri</i>	17 034	MW316634.1
叶尔羌拟双身虫 <i>P. yarkandense</i>	16 816	OM525853.1
西伯利亚拟双身虫 <i>P. homoion</i>	17 321	OM525852.1
华双身虫未定种 <i>Sindiplozoon</i> sp. DZ-2018	15 254	MG458326.1
<i>Microcotyle caudata</i>	14 267	MT180126.1
海猪多唇虫 <i>Polylabris halichoeres</i>	15 527	NC_016057.1
大西洋鲑三代虫 <i>Gyrodactylidea salaris</i>	14 790	NC_008815.1
页形指环虫 <i>Dactylogyrus lamellatus</i>	15 187	NC_035610.1
<i>O. wandeli</i>	15 185	NC_050050.1
奇异双身虫 <i>D. paradoxum</i>	15 713	PP096665

发育树于在线网页 iTOL (<https://itol.embl.de>) 中完成。

1.5 数据分析

使用 PhyloSuite^[25] 软件计算各蛋白质编码基因的核苷酸组成, 核苷酸组成偏向性计算公式^[26]: $AT-skew = [A - T] / [A + T]$ 和 $GC-skew = [G - C] / [G + C]$ 。运用 EMBOSS 在线网站 (<http://www.bioinformatics.nl/emboss-explorer/>) 的 CUSP 程序, 计算得到密码子 3 个位置的 GC 含量, 第一、二、三位上的 GC 含量分别表示为 GC1、GC2、GC3, GC12 表示 GC1 和 GC2 的平均值。利用 CodonW-1.4.2 软件 (<http://codonw.sourceforge.net>) 计算密码子组成和偏性指标, 包括相对同义密码子使用度 (relative synonymous codon usage, RSCU)、有效密码子数 (effective number of codon, ENC)。

2 结果

2.1 线粒体基因组

线粒体及蛋白质编码基因 奇异双身虫线粒体基因组测序后为线状结构 (大非编码区未测全), 其长度为 15 713 bp, 包含 36 个基因, 其中 12 个蛋白质编码基因 (PCGs)、22 个转运 RNA 基因 (tRNAs)、2 个核糖体 RNA 基因 (rRNAs), 基因

组缺少 *atp8*, 且所有基因都在同一条链 (重链) 上转录。奇异双身虫线粒体基因组排列紧密, 包含蛋白质编码基因和 tRNA 之间的重叠, 其中最大的重叠区位于 *nad4L* 和 *nad4* 之间为 49 bp, 其次是位于 *nad6* 和 *trnY* 之间的重叠为 44 bp (表 2)。

奇异双身虫线粒体基因组的蛋白质编码基因包括 5 种起始密码子: ATG、TTG、GTG、ATT 和 ATA。*nad5*、*cytb*、*nad4*、*nad4L*、*nad2*、*nad3* 和 *cox2* 起始密码子为 ATG; *atp6* 和 *nad1* 起始密码子为 TTG; *cox3*、*cox1* 和 *nad6* 起始密码子分别为 GTG、ATT 和 ATA。终止密码子包括 3 种: TAG、TAA 和 T。*nad5*、*cytb*、*nad4L*、*nad2*、*cox1* 和 *cox3* 终止密码子为 TAG; *nad6*、*nad4*、*atp6*、*nad1* 和 *cox2* 的终止密码子为 TAA; *nad3* 的终止密码子为 T (表 2)。

奇异双身虫线粒体基因组核苷酸组成为 A: 20.3%; T: 48.4%; G: 20.6%; C: 10.8%, A+T 含量为 68.7%, G+C 含量为 31.3%, 具有明显的 AT 偏向性。AT-skew 为 -0.409, GC-skew 为 0.312, 表示线粒体基因组中 T 碱基含量高于 A 碱基, G 碱基含量高于 C 碱基 (表 3)。奇异双身虫 12 个蛋白质编码基因总长为 10 272 bp, A+T 含量为 68.5%。蛋白质编码基因的第三位密码子表现出最高的 A+T 含量 (74.4%), 显著高于第一和第二位

表 2 奇异双身虫线粒体基因结构特点

Tab. 2 Characteristics of the mitochondrial genes of *D. paradoxum*

基因 gene	位置 position		大小/bp size	基因间隔 intergenic	密码子 codon		反密码子 anti-codon	链 strand
	起始 start	终止 end			起始 start	终止 end		
<i>nad6</i>	1	459	459	-44	ATA	TAA	—	H
<i>trnY</i>	416	480	65	1	—	—	GTA	H
<i>trnL1</i>	482	546	65	2	—	—	TAG	H
<i>trnS2</i>	549	604	56	2	—	—	TCA	H
<i>trnL2</i>	607	670	64	7	—	—	TAA	H
<i>trnR</i>	678	742	65	-21	—	—	CGT	H
<i>nad5</i>	722	2 254	1 533	-1	ATG	TAG	—	H
<i>trnE</i>	2 254	2 316	63	51	—	—	TTC	H
<i>cytb</i>	2 368	3 468	1 101	-17	ATG	TAG	—	H
<i>nad4L</i>	3 452	3 721	270	-49	ATG	TAG	—	H
<i>nad4</i>	3 673	4 926	1 254	196	ATG	TAA	—	H
<i>trnF</i>	5 123	5 187	65	734	—	—	TTC	H
<i>trnQ</i>	5 922	5 984	63	0	—	—	TTG	H
<i>atp6</i>	5 985	6 575	591	-17	TTG	TAA	—	H
<i>nad2</i>	6 559	7 455	897	3	ATG	TAG	—	H
<i>trnV</i>	7 459	7 520	62	9	—	—	TAC	H
<i>trnA</i>	7 530	7 592	63	-1	—	—	TGC	H
<i>trnD</i>	7 592	7 657	66	36	—	—	GTC	H
<i>nad1</i>	7 694	8 599	906	-15	TTG	TAA	—	H
<i>trnN</i>	8 585	8 648	64	3	—	—	GTT	H
<i>trnP</i>	8 652	8 714	63	7	—	—	TGG	H
<i>trnI</i>	8 722	8 788	67	1	—	—	GAT	H
<i>nad3</i>	8 790	9 063	274	-6	ATG	T	—	H
<i>trnS1</i>	9 058	9 115	58	0	—	—	AGC	H
<i>trnW</i>	9 116	9 178	63	19	—	—	TCA	H
<i>cox1</i>	9 198	10 784	1 587	-1	ATT	TAG	—	H
<i>trnG</i>	10 784	10 846	63	196	—	—	TCC	H
<i>trnT</i>	11 043	11 105	63	0	—	—	TGT	H
<i>rrnL</i>	11 106	12 073	968	0	—	—	—	H
<i>rrnS</i>	12 074	12 817	744	-1	—	—	—	H
<i>cox2</i>	12 817	13 455	639	-10	ATG	TAA	—	H
<i>trnM</i>	13 446	13 510	65	0	—	—	CAT	H
<i>trnH</i>	13 511	13 573	63	-9	—	—	GTG	H
<i>cox3</i>	13 565	14 326	762	10	GTG	TAG	—	H
<i>trnC</i>	14 337	14 401	65	2	—	—	GCA	H
<i>trnK</i>	14 404	14 466	63	586	—	—	CTT	H
<i>NCR</i>	14 467	15 713	1 247	0	—	—	—	H

注: “—”表示无。

Notes: “—” represents none.

密码子, 第三位密码子还表现出最高的 AT-skew (-0.578)(表 3)。

系统发育分析 基于本研究所测的奇异双身虫以及从 NCBI 数据库中下载的单殖吸虫的线粒体基因数据(表 1), 以 *O. wandeli* 物种作为外类群。利用线粒体基因组 12 个蛋白质编码基因构建 BI 和 ML 系统发育树。结果显示, 两种方法构建的系统发育树拓扑结构一致。进化树主要分为两个大分支, 第一大支为单独的外类群; 第二大支将寡钩亚纲 (Oligonchoinea) 和多钩亚纲 (Polychonchoinea) 的单殖吸虫聚为一支。第二大支又分为 2 个大分支, 第一大支为多钩亚纲的三代虫和指环虫, 第二大支又分为 2 个分支, 双身虫科物种聚为一支, 其余物种聚为另一支 (包含 3 科), 鲟铗虫科 (Chauhaneidae)、八铗虫科 (Diclidophoridae) 和微杯虫科 (Microcotylidae)。在双身虫科分支中, 奇异双身虫与西伯利亚拟双身虫形成姐妹支, 华双身虫未定种是这两个物种的姐妹支, 再与四个拟双身虫属物种 (云南拟双身虫、叶尔羌拟

双身虫、鳌拟双身虫、马口鱼拟双身虫) 形成姐妹群。基于上述结果, 全部双身虫科物种聚在一个大分支, 支持双身虫科物种的单系性(图 1)。

2.2 奇异双身虫线粒体基因组的密码子使用偏好

奇异双身虫密码子组成分析 按照国际通用方法, 为避免长度较短的基因带来的样本误差 (因长度过短必然会引起密码子使用偏性), 本研究只选取长度大于 300 个核苷酸的蛋白质编码基因进行分析^[27-28](除 *nad3* 基因和 *nad4L* 基因)。对奇异双身虫 10 个蛋白质编码基因进行偏好性分析(表 4), 分析得出 10 条 CDS 序列的平均 GC 含量为 31.2%(29.2%~32.8%); 密码子 3 个位置的平均 GC 含量分别为 GC1、GC2 和 GC3, GC1: 33.49%(28.11%~41.78%)、GC2: 34.06%(23.53%~39.89%) 和 GC3: 25.68%(19.72%~30.07%)。

有效密码子数 ENC 反映的是一个基因中所有用到密码子种类的多少, 其数值范围在

表 3 奇异双身虫线粒体基因组的碱基组成

Tab. 3 Nucleotide composition of mitochondrial genome in *D. paradoxum*

基因/区域 gene/regions	碱基组成/% base composition						AT偏斜 AT-skew	GC偏斜 GC-skew
	T	C	A	G	AT	GC		
全基因组 full genome	48.4	10.8	20.3	20.6	68.7	31.3	-0.409	0.312
蛋白质编码基因 PCGs	50.4	10.5	18.1	21.1	68.5	31.6	-0.472	0.336
转运RNA基因 tRNAs	38.6	12.6	26.1	22.7	64.7	35.3	-0.193	0.289
核糖体RNA基因 rRNAs	40.4	12.1	27.7	19.8	68.1	31.9	-0.186	0.239
第一位点 1st	43.2	11.3	22.4	23.1	65.6	34.4	-0.316	0.341
第二位点 2nd	49.4	12.7	16.1	21.9	65.5	34.6	-0.509	0.267
第三位点 3rd	58.7	7.4	15.7	18.2	74.4	25.6	-0.578	0.424
<i>atp6</i>	54.1	11	16.6	18.3	70.7	29.3	-0.531	0.249
<i>cox1</i>	45.2	13.5	19.5	21.7	64.7	35.2	-0.397	0.231
<i>cox2</i>	46.5	10.8	21.3	21.4	67.8	32.2	-0.372	0.33
<i>cox3</i>	49	13.4	18.2	19.4	67.2	32.8	-0.457	0.184
<i>cytb</i>	47.6	11.8	18.4	22.2	66	34	-0.442	0.305
<i>nad1</i>	52.6	9.3	16.3	21.7	68.9	31	-0.526	0.402
<i>nad2</i>	52.1	9.3	18.1	20.6	70.2	29.9	-0.485	0.381
<i>nad3</i>	54.4	6.6	16.4	22.6	70.8	29.2	-0.536	0.55
<i>nad4</i>	50	10	18.7	21.2	68.7	31.2	-0.455	0.357
<i>nad4L</i>	53	6.3	21.9	18.9	74.9	25.2	-0.416	0.5
<i>nad5</i>	54.3	8.1	16.1	21.5	70.4	29.6	-0.543	0.453
<i>nad6</i>	54.9	8.9	16.1	20	71	28.9	-0.546	0.383
<i>rrnL</i>	41.9	11.2	26.8	20.1	68.7	31.3	-0.221	0.287
<i>rrnS</i>	38.3	13.4	28.9	19.4	67.2	32.8	-0.14	0.18

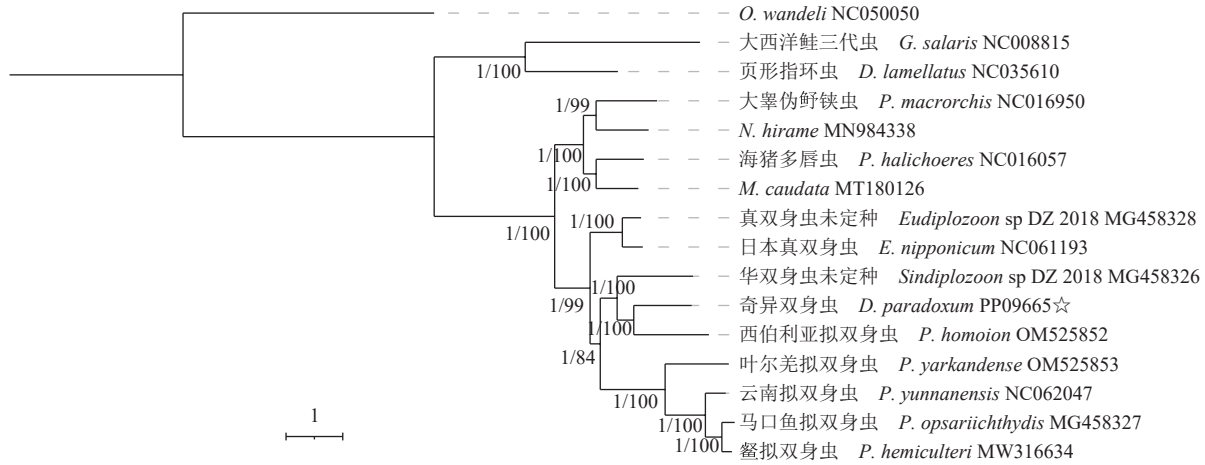


图 1 基于蛋白质编码基因和贝叶斯法、最大似然法构建的单殖吸虫系统发育树

左侧为进化树, 物种名称后跟 GenBank 登录号, 分支节点数值表示自展检验支持率。

Fig. 1 Phylogenetic tree of monogenean based on protein-coding genes and constructed by Bayesian inference and Maximum likelihood methods

The left side is the phylogenetic tree, the species name followed by the GenBank accession numbers, the branch node values represent the support rate for self expansion test.

表 4 奇异双身虫线粒体基因组不同位置的 GC 含量及 ENC 值

Tab. 4 GC content and ENC value of different regions in mitochondrial genome of *D. paradoxum*

基因 gene	GC1/%	GC2/%	GC3/%	GC/%	ENC
<i>atp6</i>	28.11	33.51	25.95	29.4	32.15
<i>cox1</i>	40.83	39.89	24.95	32.4	34.6
<i>cox2</i>	41.78	35.21	19.72	32.4	33.87
<i>cox3</i>	37.4	36.61	24.41	32.8	39.37
<i>cytb</i>	36.51	35.97	29.43	34	38.06
<i>nad1</i>	32.12	36.09	24.83	31.1	34.48
<i>nad2</i>	31.44	33.78	24.41	29.9	37.04
<i>nad4</i>	32.54	32.54	28.71	31.3	43.2
<i>nad5</i>	30.92	33.46	24.27	29.5	34.92
<i>nad6</i>	33.33	23.53	30.07	29.2	34.22

注: GC1, GC2, GC3和GC分别代表密码子第1, 2, 3位和总的GC含量; ENC. 有效密码子数。

Notes: GC1, GC2, GC3 and GC represent the first, second, third and total GC content, respectively; ENC. Effective number of codons.

20~61, ENC 值越小, 密码子使用偏好越强, 值得注意的是 ENC 值受碱基组成影响^[29]。对奇异双身虫 10 条蛋白质编码基因序列的 ENC 值进行统计, ENC 取值^[30]35 作为区分密码子偏好强弱的标准 (ENC 范围 20~61)(表 4), 最终得出所有基因的 ENC 范围为 32.15~43.2, *atp6*、*cox1*、*cox2*、*nad1*、*nad5* 和 *nad6* 基因的 ENC 平均值均小于 35, 密码子偏好性较强; *cox3*、*cytb*、*nad2*、和 *nad4* 基因的 ENC 平均值均大于 35, 密码子偏好性较弱。

2.3 双身虫科物种密码子偏好性

双身虫科物种密码子组成分析 对双身虫科物种蛋白质编码基因的 ENC 平均值进行统计分析, 得出 9 个双身虫科物种的 ENC 范围为 30.632~37.495(图 2)。其中日本真双身虫、叶尔羌拟双身虫、云南拟双身虫和华双身虫未定种的 ENC 平均值分别为 33.229、32.154、30.632 和 33.937

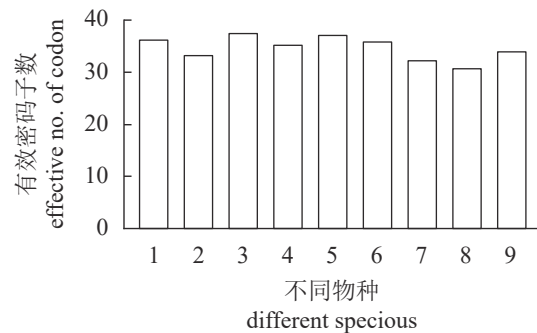


图 2 9 种双身虫物种的线粒体蛋白质编码基因 ENC 平均值比较分析

1. 奇异双身虫; 2. 日本真双身虫; 3. 真双身虫未定种; 4. 蟹拟双身虫;
5. 西伯利亚拟双身虫; 6. 马口鱼拟双身虫; 7. 叶尔羌拟双身虫;
8. 云南拟双身虫; 9. 华双身虫未定种。

Fig. 2 Comparative analysis of the mean values of mitochondrial protein-coding genes ENC in nine species of Diplozoidae

1. *D. paradoxum* 2. *E. nipponicum* 3. *Eudiplozoon*. sp. DZ-2018 4. *P. hemiculteri* 5. *P. homoion* 6. *P. opsariichthydis* 7. *P. yarkandense* 8. *P. yunnanensis* 9. *Sindiplozoon* sp. DZ-2018.

均小于 35, 密码子偏好性较强。奇异双身虫、真双身虫未定种、鳖拟双身虫、西伯利亚拟双身虫和马口鱼拟双身虫的 ENC 平均值分别为 36.191、37.495、35.271、37.163 和 35.821 均大于 35, 密码子偏好性较弱。上述结果表明, 双身虫科不同物种线粒体蛋白质编码基因的密码子使用偏好的强弱存在差异。

对 9 种双身虫的 GC、ENC 值、密码子数目 N 进行相关性分析(表 5)。真双身虫未定种、马口鱼拟双身虫和云南拟双身虫的 GC1 与 GC2 显著相关。双身虫科物种 GC 与 GC1 和 GC2 均有相关性(除西伯利亚拟双身虫 GC 仅与 GC1 有相关性), 而 GC1、GC2 与 GC3 均不显著相关。说明大多双身虫物种的线粒体基因组密码子的第一、第二位碱基相似度高, 与第三位碱基差别较大。叶尔羌拟双身虫的 ENC 与 GC3 呈极显著相关, 说明叶尔羌拟双身虫第三位碱基对密码子偏好影响较大。日本真双身虫的密码子数 N 与 GC3 的相关性呈显著水平, 其余双身虫科物种两者相关性不显著。

相对同义密码子使用度 (RSCU) 分析

RSCU 是指对于某一特定的密码子在编码对应氨基酸的同义密码子间的相对概率, 常用来研究基因中全部同义密码子偏好变化^[31]。RSCU 值大于 1 的密码子表现出强偏好(即使用频率高于其他同义密码子); RSCU 值小于 1 的密码子表现出弱偏好(即使用频率低于其他同义密码子); 而 RSCU 值等于 1 的密码子不具偏好性^[32]。对获得的蛋白质编码基因的 RSCU 值进行分析(图 3), 结果表明 9 个双身虫科物种蛋白质编码基因中 UUU、UUA、UUG、AUU、GUU、UCU、CCU、ACU、GCU、UAU、CAU、AAU、GAU、GAG、UGU、CGU、AGU 和 GGU 等 18 个密码子的 RSCU 值都大于 1, 是 9 个双身虫科物种共同偏好的密码子即高频密码子。日本真双身虫的 UCA 和 ACA 两个密码子 RSCU 值大于 1, 而在其他 8 个双身虫科物种中均小于 1, 故 UCA 和 ACA 为日本真双身虫特有的偏好性密码子。同样, ACG、GCC 和 AGA 为叶尔羌拟双身虫特有的偏好性密码子; ACC、UCC 和 UGA 分别为真双身虫未定种、马口鱼拟双身虫和云南拟双身虫特有的偏好性密码子。综上, 9 种双身虫科物种线粒体基因组蛋白质编码基因都偏好使用以 U 结尾的密码子, 避免使用以 C 结尾的密码子。

PR2-plot 分析 PR2-plot 通过分析密码子第 3 位碱基的组成关系, 揭示影响密码子使用偏好的影响因素, 如果只受到突变单一因素影响时, 4 种碱基会被均衡使用。以各基因的 A3/(A3+T3) 和 G3/(G3+C3) 为纵坐标和横坐标, (0.5, 0.5) 坐标为中心绘图, 从这个中心出发的矢量代表了奇偶偏好的方向和程度^[33-34]。双身虫科物种的 PR2-plot 绘图分析结果表明(图 4), 双身虫科物种的全部基因都分布于平面图的右下方, 说明密码子最后一位碱基的使用具有偏好性, T 的使用频率高于 A, G 的使用频率高于 C。说明突变不是影响双身虫科物种的密码子使用偏好性的唯一因素, 选择等其他因素对其密码子使用偏好也产生一定的影响。

中性绘图分析 中性绘图分析(neutrality plot)是一种衡量密码子偏好形成原因的有效方法, 可以很容易地检测出形成密码子使用偏好的主要力量^[35]。中性绘图分析是以 GC12 为纵坐标, GC3 为横坐标^[36], 通过分析 GC12 与 GC3 之间的相关性来研究密码子偏好的影响因素。如果图中回归的斜率非常接近 1, 表明密码子 3 个位置的碱基组成无差异, 突变是影响密码子的偏好的主要力量。如果斜率接近 0, 表明密码子 3 个位置的碱基组成存在差异, 自然选择是形成密码子偏好的主要力量^[36]。双身虫科物种的 GC12 和 GC3 之间的中性绘图分析(图 5)表明日本真双身虫、真双身虫未定种、鳖拟双身虫、西伯利亚拟双身虫、叶尔羌拟双身虫、云南拟双身虫和华双身虫未定种的标准曲线的斜率分别为 0.076 89、0.432 4、0.319 6、0.412 7、0.139 2、0.131 1 和 0.165 1。表示突变压力的影响分别为 7.689%、43.24%、31.96%、41.27%、13.92%、13.11% 和 16.51%, 自然选择的影响分别为 92.311%、56.76%、68.04%、58.73%、86.08%、86.89% 和 83.49%。说明自然选择比突变压力对这些物种的密码子使用偏好影响大。而奇异双身虫和马口鱼拟双身虫的标准曲线的斜率分别是 0.618 9 和 0.513 1。表示突变压力的影响分别为 61.89% 和 51.31%, 自然选择的影响分别为 38.11% 和 48.69%, 说明突变压力比自然选择对两个物种的密码子使用偏好影响大。综上, 9 个双身虫科物种密码子使用偏好受突变压力或自然选择的强度存在差异。同时, 9 个双身虫科物种的 GC12 与 GC3 含量有的呈正相关有的呈负相关, 但相关性都不显著, 进一步表明密码子第三位碱基与第一、二位碱基组成不相同。

表 5 9 种双身虫线粒体基因组密码子偏好性参数的相关性分析结果

Tab. 5 Correlation analysis of codon usage bias parameters of the mitochondrial genomes of nine species of Diplozoidae

物种 species	参数 parameter	GC1	GC2	GC3	GC	ENC
奇异双身虫 <i>D. paradoxum</i>	GC2	0.429				
	GC3	-0.373	-0.499			
	GC	0.753*	0.638*	-0.095		
	ENC	0.049	0.073	0.351	0.382	
	N	-0.309	0.16	0.092	0.01	0.483
日本真双身虫 <i>E. nipponicum</i>	GC2	0.601				
	GC3	-0.08	-0.012			
	GC	0.85**	0.84**	0.269		
	ENC	-0.02	0.262	0.401	0.269	
	N	-0.441	-0.011	-0.665*	-0.485	0.015
真双身虫未定种 <i>Eudiplozoon</i> . sp. DZ-2018	GC2	0.665*				
	GC3	0.149	0.455			
	GC	0.834**	0.927**	0.444		
	ENC	0.029	0.03	0.275	0.181	
	N	-0.261	-0.114	-0.297	-0.092	0.241
蚤拟双身虫 <i>P. hemiculteri</i>	GC2	0.537				
	GC3	0.262	0.008			
	GC	0.91**	0.766**	0.421		
	ENC	0.095	-0.099	0.575	0.132	
	N	-0.249	-0.042	0.439	-0.05	-0.176
西伯利亚拟双身虫 <i>P. homoion</i>	GC2	0.309				
	GC3	-0.15	-0.745*			
	GC	0.838**	0.201	0.205		
	ENC	-0.026	0.149	0.116	0.179	
	N	-0.426	0.442	-0.24	-0.159	0.419
马口鱼拟双身虫 <i>P. opsariichthydis</i>	GC2	0.635*				
	GC3	-0.154	-0.418			
	GC	0.858**	0.762*	0.081		
	ENC	-0.408	-0.434	-0.147	-0.565	
	N	-0.384	-0.028	0.366	0.038	-0.26
叶尔羌拟双身虫 <i>P. yarkandense</i>	GC2	0.449				
	GC3	-0.191	0.062			
	GC	0.805**	0.789**	0.26		
	ENC	0.031	0.35	0.776**	0.471	
	N	-0.53	-0.253	0.058	-0.449	-0.21
云南拟双身虫 <i>P. yunnanensis</i>	GC2	0.671*				
	GC3	0.246	-0.144			
	GC	0.941**	0.780**	0.313		
	ENC	-0.316	-0.412	0.432	-0.29	
	N	-0.398	0.053	-0.252	-0.197	-0.107
华双身虫未定种 <i>Sindiplozoon</i> . sp. DZ-2018	GC2	0.374				
	GC3	0.112	0.117			
	GC	0.733*	0.654*	0.608		
	ENC	-0.116	0.362	0.256	0.284	
	N	-0.429	0.328	-0.068	-0.026	0.443

注: *表示显著相关($P<0.05$); **表示极显著相关($P<0.01$)。Notes: * indicates significant correlation ($P<0.05$); ** indicates extremely significant correlation ($P<0.01$).

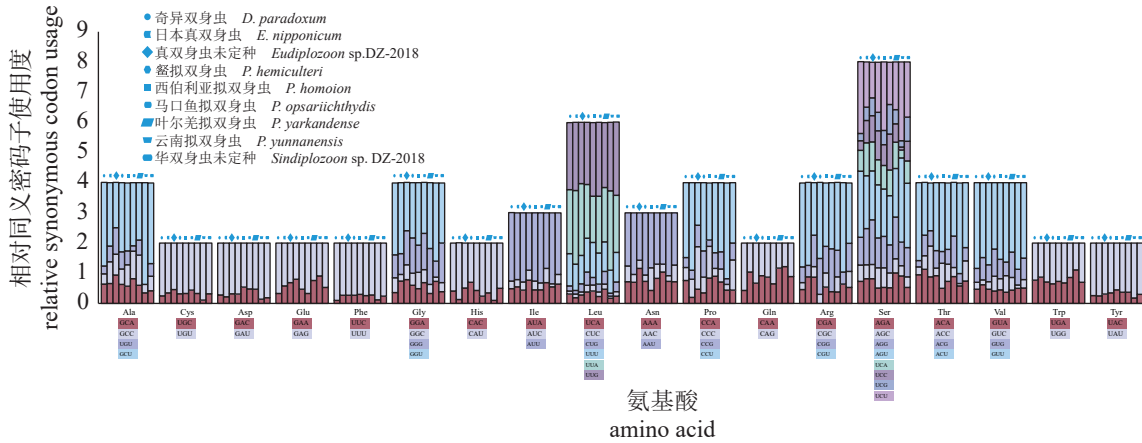


图3 双身虫科线粒体蛋白质编码基因相对同义密码子使用情况

Fig. 3 Relative synonymous codon usage in mitochondrial protein-coding genes of Diplozoidae

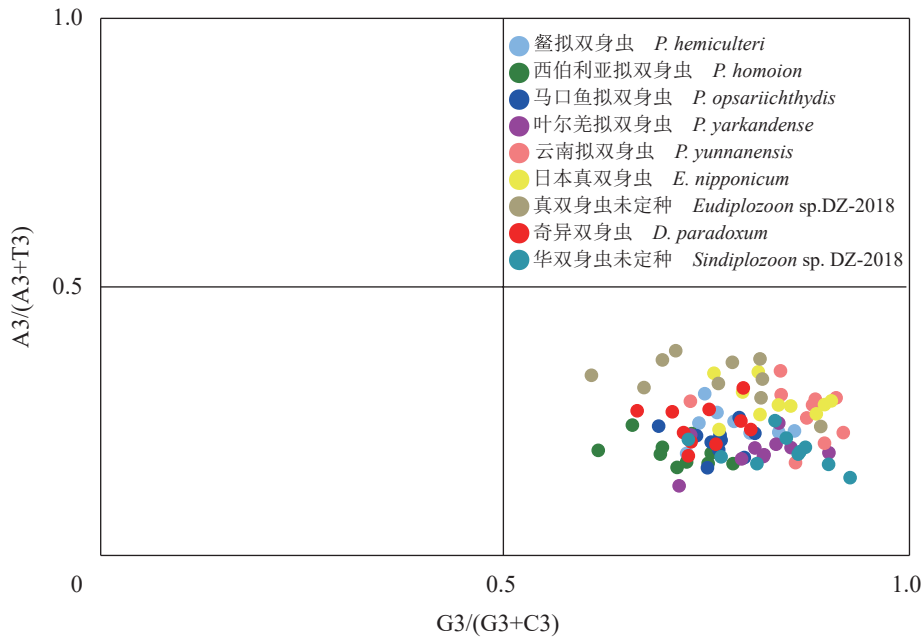


图4 9个双身虫科物种线粒体基因组的PR2-plot分析图

Fig. 4 PR2-plot analysis of mitochondrial genomes of nine species of Diplozoidae

3 讨论

对双身虫科物种密码子使用偏好及偏好性影响因素分析, 通过密码子组成及RSCU分析得到它们都受到碱基组成的影响。通过PR2-plot分析及中性绘图分析得到日本真双身虫、真双身虫未定种、螯拟双身虫、西伯利亚拟双身虫、叶尔羌拟双身虫、云南拟双身虫和华双身虫未定种受突变压力比自然选择影响小; 奇异双身虫和马口鱼拟双身虫受突变压力比自然选择影响大。综上, 9种双身虫科物种受到突变压力、自然选择和碱基组成的共同影响。

奇异双身虫线粒体基因组蛋白质编码基因的AT含量为68.5%, 其他8种双身虫线粒体基因组蛋白编码基因AT含量在65.2%~73.4%, 密码子都偏好以A/U结尾。本研究结果与其他物种的结果相似, 如杨兴等^[37]在研究亚洲带绦虫(*Taenia asiatica*)和范俊杰等^[38]在研究细粒棘球绦虫(*Echinococcus granulosus*)的碱基组成时, 发现二者编码区GC含量均高于50%, 且密码子都偏好以G/C结尾。Mazumder等^[39]在对扁形动物ND基因的碱基组成分析时, 也指出扁形动物的基因组AT含量高于GC含量, 且密码子大都以A/U结尾。以上研究表明, GC含量高的物种密码子更偏好以

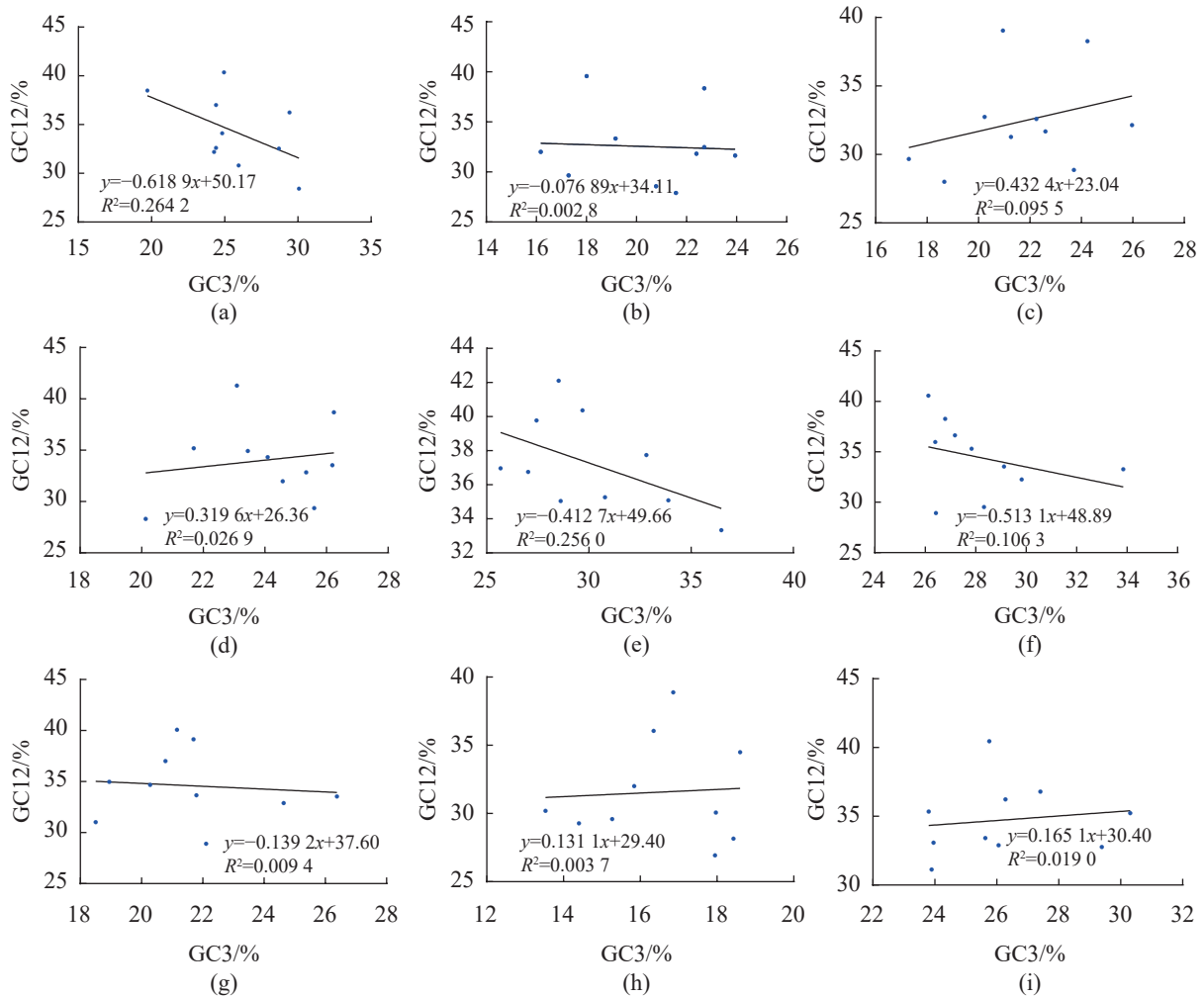


图 5 9种双身虫物种线粒体基因组中性绘图分析结果

(a) 奇异双身虫, (b) 日本真双身虫, (c) 真双身虫未定种, (d) 鳖拟双身虫, (e) 西伯利亚拟双身虫, (f) 马口鱼拟双身虫, (g) 叶尔羌拟双身虫, (h) 云南拟双身虫, (i) 华双身虫未定种。

Fig. 5 Neutrality plot analysis of mitochondrial genomes of nine species of Diplozoidae

(a) *D. paradoxum*, (b) *E. nipponicum*, (c) *Eudiplozoon* sp. DZ-2018, (d) *P. hemiculteri*, (e) *P. homoion*, (f) *P. opsariichthydis*, (g) *P. yarkandense*, (h) *P. yunnanensis*, (i) *Sindiplozoon* sp. DZ-2018.

G/C 结尾, AT 含量高的物种密码子更偏好以 A/U 结尾。由此可知奇异双身虫及双身虫科物种的密码子使用偏好与碱基组成有着密不可分的关系, 其不但是线粒体基因组的一个重要特征, 也是影响密码子使用偏好的一个重要因素。

通过 PR2-plot 分析可知, 影响密码子偏好性的因素是随机突变(此时密码子 A3≠T3, G3≠C3)。本研究结果显示, 双身虫科物种的 A、T、C、G 呈不均匀分布(图 4), 且 T 和 G 的使用频率高于 A 和 C, 表明双身虫科物种的线粒体蛋白质编码基因同时受到了突变压力和自然选择等其他因素的影响。与 Mazumder 等^[39]对多种扁形动物包括

绦虫纲 (Cestoda)、单殖吸虫纲、复殖吸虫 (Rabd- itophora)、吸虫纲 (Trematoda)、涡虫纲 (Turbel- laria) 的 ND 基因进行 PR2 分析结果一致, AT 和 GC 碱基的使用不平等, T 和 G 的使用频率高于 A 和 C。而与杨兴等^[37]和黄兴^[40]对亚洲带绦虫和多头带绦虫 (*T. multiceps*) 的 PR2 分析结果不同, T 和 C 使用频率高于 A 和 G。以上研究均指出除突变压力外, 其他的因素(如自然选择等)也会影响扁形动物的使用。综上可知扁形动物的密码子偏好性不仅受突变压力的作用, 自然选择等其他因素都影响偏好性的形成。

通过中性绘图以 GC12 与 GC3 之间的相关性来估计突变压力和自然选择对密码子使用的偏好程度^[36], 本研究指出奇异双身虫和马口鱼拟双身

虫受突变压力的影响比自然选择影响大; 日本真双身虫、真双身虫未定种、蟹拟双身虫、西伯利亚拟双身虫、叶尔羌拟双身虫、云南拟双身虫和华双身虫未定种受突变压力的影响比自然选择影响小。由此可知, 双身虫科物种的密码子使用偏好的影响强弱不一致, 但大多数受到自然选择的影响较大。本研究大多双身虫物种结果与其他扁形动物的结果一致, 如 Mazumder 等^[41] 在 2021 年对扁形动物的 3 个细胞色素 c 氧化酶亚基偏好性进行研究, 中性绘图分析表明突变压力的影响比自然选择的影响小。杨兴等^[37] 在对 6 种绦虫进行中性绘图分析时, 发现 6 种绦虫受到突变压力的影响比自然选择影响小。黄兴^[40] 等在对多头带绦虫进行中性绘图分析时, 也指出多头带绦虫受到突变压力的影响比自然选择影响小。以上研究均指出大多扁形动物的密码子使用偏好都受到突变压力和自然选择的影响, 且受自然选择的影响更强。

线粒体蛋白质编码基因密码子使用偏好在双身虫科中表现出略微差异性, 造成这种差异的原因可能与物种的分布范围广及宿主的生存环境有关。双身虫科物种分布范围广, 在中国西北、中部、西南等区域都有研究发现, 且双身虫科物种寄生多种鱼类的生存环境不同。如奇异双身虫寄生于东方欧鳊属于广温性鱼类, 生活在温带大陆性干旱半干旱气候中国西北的额尔齐斯河流域。叶尔羌拟双身虫寄生于斑重唇鱼 (*Diptychus maculatus*) 属于冷水性鱼类, 生活在高海拔叶尔羌河流域。云南拟双身虫寄生于短吻鱼 (*Sikukia stejnegeri*)、黄尾短吻鱼 (*Sikukia flavicaudata*), 生活在低海拔的云南省澜沧江。本研究结果与姜艳^[42] 的研究相似, 研究指出小鲢科 (Hynobiidae) 不同物种的密码子偏好性主要影响因素存在差异, 猜测与小鲢科各物种分布范围广, 个体跨越栖息地的能力非常弱, 各物种生存环境不同, 承受压力不同, 突变和选择在物种之间有不同程度的差异有关。总体而言, 虽然双身虫科密码子偏好性存在略微差异, 但都受突变压力、自然选择和碱基组成的共同影响。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

[1] 王新, 焦丽, 姚卫建, 等. 额尔齐斯河鱼类双身虫科一

种属新记录 [J]. 水生生物学报, 2015, 39(4): 794-797.

Wang X, Jiao L, Yao W J, et al. The first record of a newly reported Chinese diplozoidae of Monogenea on fishes in the Ergis river[J]. Acta Hydrobiologica Sinica, 2015, 39(4): 794-797 (in Chinese).

[2] 王鹏飞, 高欢欢, 苏玲, 等. 2 种果蝇 (*Drosophila melanogaster* 与 *D. sechellia*) 线粒体及 NADH dehydrogenase subunit 基因的密码子偏好性分析 [J]. 安徽农业科学, 2018, 46(1): 96-101.

Wang P F, Gao H H, Su L, et al. Comprehensive analysis of codon usage bias of *Drosophila melanogaster* and *D. sechellia* mitochondrion and NADH dehydrogenase subunit gene family[J]. Journal of Anhui Agricultural Sciences, 2018, 46(1): 96-101 (in Chinese).

[3] 武志娟, 钟金城. 密码子偏性及其应用 [J]. 生物学通报, 2012, 47(4): 9-11.

Wu Z J, Zhong J C. Codon bias and its application[J]. Bulletin of Biology, 2012, 47(4): 9-11 (in Chinese).

[4] Bulmer M. The selection-mutation-drift theory of synonymous codon usage[J]. Genetics, 1991, 129(3): 897-907.

[5] 冯瑞云, 梅超, 王慧杰, 等. 籽粒苋叶绿体基因组密码子偏好性分析 [J]. 中国草地学报, 2019, 41(4): 8-15.

Feng R Y, Mei C, Wang H J, et al. Analysis of codon usage in the chloroplast genome of grain amaranth (*Amaranthus hypochondriacus* L.) [J]. Chinese Journal of Grassland, 2019, 41(4): 8-15 (in Chinese).

[6] 任桂萍, 董璿莹, 党云琨. 密码子中的密码: 密码子偏好性与基因表达的精细调控 [J]. 中国科学: 生命科学, 2019, 49(7): 839-847.

Ren G P, Dong Y Y, Dang Y K. Codon codes: codon usage bias influences many levels of gene expression[J]. Scientia Sinica Vitae, 2019, 49(7): 839-847 (in Chinese).

[7] 原晓龙, 康洪梅, 王毅. 云南蓝果树叶绿体基因组密码子偏好性分析 [J]. 西北林学院学报, 2020, 35(4): 26-31, 124.

Yuan X L, Kang H M, Wang Y. Codon usage bias analysis of chloroplast genome in *Nyssa yunnanensis*[J]. Journal of Northwest Forestry University, 2020, 35(4): 26-31, 124 (in Chinese).

[8] Sharp P M, Emery L R, Zeng K. Forces that influence the evolution of codon bias[J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2010, 365(1544): 1203-1212.

- [9] Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*[J]. Proceedings of the National Academy of Sciences of the United States of America, 1999, 96(8): 4482-4487.
- [10] Comeron J M, Kreitman M, Aguadé M. Natural selection on synonymous sites is correlated with gene length and recombination in drosophila[J]. *Genetics*, 1999, 151(1): 239-249.
- [11] 周丹, 薛仁余, 张晓峰, 等. 鲤和斑马鱼 *HOX* 基因家族同义密码子使用偏性的分析 [J]. *水产学杂志*, 2013, 26(2): 19-25.
- Zhou D, Xue R Y, Zhang X F, *et al.* Synonymous codon bias of HOX gene family in common carp (*Cyprinus carpio*) and zebrafish (*Danio rerio*)[J]. *Chinese Journal of Fisheries*, 2013, 26(2): 19-25 (in Chinese).
- [12] Vinogradov A E. Intron length and codon usage[J]. *Journal of Molecular Evolution*, 2001, 52(1): 2-5.
- [13] Goodarzi H, Torabi N, Najafabadi H S, *et al.* Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons[J]. *Gene*, 2008, 407(1-2): 30-41.
- [14] Berg O G. Selection intensity for codon bias and the effective population size of *Escherichia coli*[J]. *Genetics*, 1996, 142(4): 1379-1382.
- [15] 吴炜倩, 胡源, 胡接力, 等. 乙型肝炎病毒密码子偏好性及影响因素分析 [J]. *中国病原生物学杂志*, 2020, 15(12): 1404-1410.
- Wu W Q, Hu Y, Hu J L, *et al.* Codon usage bias of the hepatitis B virus and factors influencing it[J]. *Journal of Pathogen Biology*, 2020, 15(12): 1404-1410 (in Chinese).
- [16] Hershberg R, Petrov D A. Selection on codon bias[J]. *Annual Review of Genetics*, 2008, 42(1): 287-299.
- [17] 王新, 朱梦莹, 周颖, 等. 额尔齐斯河东方欧鳊寄生奇异双身虫的种群生态学研究 [J]. *水生态学杂志*, 2014, 35(4): 87-91.
- Wang X, Zhu M Y, Zhou Y, *et al.* Population ecology of *Diplozoon paradoxum* Nordmann in *Abramis brama orientalis* berg in Ergis river[J]. *Journal of Hydroecology*, 2014, 35(4): 87-91 (in Chinese).
- [18] Coil D, Jospin G, Darling A E. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data[J]. *Bioinformatics*, 2015, 31(4): 587-589.
- [19] Bankevich A, Nurk S, Antipov D, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing[J]. *Journal of Computational Biology*, 2012, 19(5): 455-477.
- [20] Walker B J, Abeel T, Shea T, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement[J]. *PLoS One*, 2014, 9(11): e112963.
- [21] Bernt M, Donath A, Jühling F, *et al.* MITOS: improved *de novo* metazoan mitochondrial genome annotation[J]. *Molecular Phylogenetics and Evolution*, 2013, 69(2): 313-319.
- [22] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Molecular Biology and Evolution*, 2013, 30(4): 772-780.
- [23] Kalyaanamoorthy S, Minh B Q, Wong T K F, *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates[J]. *Nature Methods*, 2017, 14(6): 587-589.
- [24] Nguyen L T, Schmidt H A, Von Haeseler A, *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Molecular Biology and Evolution*, 2015, 32(1): 268-274.
- [25] Zhang D, Gao F L, Jakovlić I, *et al.* PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies[J]. *Molecular Ecology Resources*, 2020, 20(1): 348-355.
- [26] Perna N T, Kocher T D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes[J]. *Journal of Molecular Evolution*, 1995, 41(3): 353-358.
- [27] Gupta S K, Bhattacharyya T K, Ghosh T C. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection[J]. *Journal of Biomolecular Structure and Dynamics*, 2004, 21(4): 527-535.
- [28] Peixoto L, Zavala A, Romero H, *et al.* The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium meliloti*[J]. *Gene*, 2003, 320: 109-116.
- [29] Jia X, Liu S Y, Zheng H, *et al.* Non-uniqueness of factors constraint on the codon usage in *Bombyx mori*[J]. *BMC Genomics*, 2015, 16(1): 356.

- [30] Jiang Y, Deng F, Wang H L, *et al.* An extensive analysis on the global codon usage pattern of baculoviruses[J]. *Archives of Virology*, 2008, 153(12): 2273-2282.
- [31] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms[J]. *Journal of Molecular Evolution*, 1986, 24(1-2): 28-38.
- [32] Villada J C, Brustolini O J B, Da Silveira W B. Integrated analysis of individual codon contribution to protein biosynthesis reveals a new approach to improving the basis of rational gene design[J]. *DNA Research*, 2017, 24(4): 419-434.
- [33] Sueoka N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons[J]. *Journal of Molecular Evolution*, 1995, 40(3): 318-325.
- [34] Sueoka N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses[J]. *Journal of Molecular Evolution*, 2001, 53(4-5): 469-476.
- [35] 原晓龙, 李云琴, 王毅, 等. 西南桦叶绿体基因组密码子偏好性分析 [J]. 基因组学与应用生物学, 2020, 39(12): 5758-5764.
- Yuan X L, Li Y Q, Wang Y, *et al.* Analysis of codon usage in the chloroplast genome of *Betula alnoides*[J]. *Genomics and Applied Biology*, 2020, 39(12): 5758-5764 (in Chinese).
- [36] Sueoka N. Directional mutation pressure and neutral molecular evolution[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1988, 85(8): 2653-2657.
- [37] 杨兴, 刘西, 盛梦诗, 等. 亚洲带绦虫基因组密码子使用偏性分析 [J]. *动物医学进展*, 2021, 42(7): 57-63.
- Yang X, Liu X, Sheng M S, *et al.* Analysis of codon usage bias in *Taenia asiatica* genome[J]. *Progress in Veterinary Medicine*, 2021, 42(7): 57-63 (in Chinese).
- [38] 范俊杰, 吴宏焯, 李锴, 等. 细粒棘球绦虫水通道蛋白基因家族成员密码子偏好性分析 [J]. 基因组学与应用生物学, 2020, 39(6): 2832-2841.
- Fan J J, Wu H Y, Li K, *et al.* The codon bias analysis for aquaporin gene family members from *Echinococcus granulosus*[J]. *Genomics and Applied Biology*, 2020, 39(6): 2832-2841 (in Chinese).
- [39] Mazumder G A, Uddin A, Chakraborty S. Analysis of codon usage pattern of mitochondrial ND genes in Platyhelminthes[J]. *Molecular and Biochemical Parasitology*, 2020, 238: 111294.
- [40] 黄兴. 多头带绦虫功能基因的原核表达及蛋白特性分析 [D]. 雅安: 四川农业大学, 2016.
- Huang X. Prokaryotic expression and recombinant protein analysis of functional genes from *Taenia multiceps*[D]. Ya'an: Sichuan Agricultural University, 2016 (in Chinese).
- [41] Mazumder G A, Uddin A, Chakraborty S. Analysis of codon usage bias in mitochondrial CO gene among Platyhelminthes[J]. *Molecular and Biochemical Parasitology*, 2021, 245: 111410.
- [42] 姜艳. 小颚科线粒体基因组密码子偏好性及进化关系的分析 [D]. 西安: 陕西师范大学, 2016.
- Jiang Y. Analysis of mitochondrial genome codon preference and its evolutionary relationship in Hynobiidae[D]. Xi'an: Shaanxi Normal University, 2016 (in Chinese).

Determination of mitochondrial genome and analysis of codon usage bias of *Diplozoon paradoxum*

SHI Caixia¹, WANG Jinpu¹, ZHANG Li¹, LIU Yanjun¹,
WEI Nianwen¹, YUE Cheng¹, HAO Cuilan^{1,2*}

(1. College of Veterinary Medicine, Xinjiang Agricultural University, Urumqi 830052, China;

2. Xinjiang Key Laboratory of New Drug Study and Creation for Herbivorous Animals, Urumqi 830052, China)

Abstract: To determine the basic structural features of the mitochondrial genome, explore the taxonomic status of *Diplozoon paradoxum*, and to understand the codon usage bias in Diplozoidae species, the mitochondrial genome sequence of *D. paradoxum* was obtained by the second generation high-throughput sequencing technologies, annotated by MITOS Web Server, and its structure was analyzed by PhyloSuite (version 1.2.2). Then, we analyzed the codon composition and usage bias of mitochondrial protein-coding genes in nine species of Diplozoidae. The results showed that the mitochondrial genome sequence of *D. paradoxum* was 15,713 bp in length (the large non-coding region was not fully sequenced), and the AT content of *D. paradoxum* was 68.7%, which had an obvious AT bias. Based on the available data, the results of phylogenetic tree analysis showed that all species of Diplozoidae were grouped into one branch, and revealed that Diplozoidae was a monophyletic group. The analysis of codon usage bias showed that the effective number of codons (ENC) ranged from 30.632 to 37.495 in Diplozoidae. The relative synonymous codon usage (RSCU) values of 18 codons were > 1 in Diplozoidae, and there was a bias for codons ending in U(T). The analysis of PR2-plot showed that the frequency of T was higher than that of A, and the frequency of G was higher than that of C. The codon usage bias might be affected by factors such as mutation and selection. The analysis of neutral mapping showed that the effects of mutation pressure on *D. paradoxum* and *P. opsariichthydis* were 61.89% and 51.31%, respectively, while the effects of mutation pressure on other seven species were below 50%. Diplozoidae species had slight differences in codon usage bias, but they were all affected by mutation pressure, natural selection, and base composition. This study can provide an important basis for follow-up phylogenetic studies.

Key words: *Diplozoon paradoxum*; mitochondrial genome; codon preference; phylogenesis

Corresponding author: HAO Cuilan. E-mail: haocuilan@126.com

Funding projects: The Key Research and Development Program for Special Project of Xinjiang Uygur Autonomous Region (2023B02012); National Natural Science Foundation of China (31960737); Natural Science Foundation of Xinjiang Uygur Autonomous Region (2021D01B56)