

# 贝叶斯联合模型与中国洞穴鱼类分化时间的估算 (鲤科: 金线鲃属)

李志强, 郭宝成, 李俊兵, 何舜平\*, 陈宜瑜

中国科学院水生生物研究所, 武汉 430072;

中国科学院研究生院, 北京 100049

\* 联系人, E-mail: [clad@ihb.ac.cn](mailto:clad@ihb.ac.cn)

2008-01-07 收稿, 2008-05-04 接受

国家自然科学基金重点项目(批准号: 30530120)和国家重点基础研究发展计划(批准号: 2007CB411600)资助

**摘要** 金线鲃属分布于中国云贵高原及周边地区, 包括 10 多种适应于洞穴生活、具有不同程度眼睛和色素退化的典型洞穴鱼类。迄今为止, 金线鲃属的系统发育关系还存在很大争议。我们获得了 34 种金线鲃属鱼类及外类群的细胞色素 *b* 和 ND4 基因序列, 用贝叶斯联合模型分析数据, 重建了金线鲃属的系统发育关系; 并用松散分子钟估算了本属及内部分支的起源时间。金线鲃属的单系性得到强烈的支持, 本属可分为 6 个分支。此外, 本研究结果还支持洞穴物种的多系起源假说以及季氏金线鲃(*Sinocyclocheilus jii*)的基部位置。松散分子钟估算表明, 金线鲃属起源于晚中新世, 在 11 百万年前(million years ago, Ma)左右, 比以前设想的要古老。

**关键词**  
金线鲃属  
系统发育  
松散分子钟

洞穴生物由于其奇特的特征, 如有些感觉器官和附肢的加强、眼睛和色素的退化或消失, 引起了人们极大关注<sup>[1,2]</sup>。洞穴生物可作为研究自然选择及适应性进化的模式系统<sup>[1]</sup>。目前, 已有研究利用墨西哥脂鲤(*Astyanax mexicanus*)作为模式生物来研究洞穴生物及其洞穴特征的演化<sup>[3-5]</sup>。在硬骨鱼类中, 除了墨西哥脂鲤外, 金线鲃属(*Sinocyclocheilus*)也可作为另外一个模型, 因为此属包含了众多洞穴物种, 并且这些物种具有不同程度的适应洞穴生活的特征, 包括眼睛和色素不同程度的退化及向前突出的角状结构(称为头角或顶骨突, 通常由顶骨后部或上枕骨共同参与形成)。金线鲃属是中国特有鱼类, 隶属于鲤科 亚科, 俗名金线鱼, 分布局限于云贵高原喀斯特岩溶地貌发育地区的溶洞及地表河流和湖泊, 包括云南省东部、贵州省南部及广西壮族自治区北部。至 2004 年, 本属已描述的有效种共为 51 种(53 个指名种)<sup>[6]</sup>。

如果用金线鲃属作为模型来研究洞穴生物的自然选择及适应性进化, 首先要明确其属内物种的系统发育关系。虽然对金线鲃属的系统发育做了很多

工作, 但其属内关系仍存在较大的争议<sup>[7-9]</sup>, 尤其是属内的洞穴物种。单乡红和乐佩琦<sup>[7]</sup>的形态学结果显示, 洞穴物种是二分支的; 但是王大忠等人基于 28 个形态及骨骼性状认为洞穴物种是一单系群<sup>[8]</sup>。金线鲃属洞穴物种的一些形态或骨骼性状是与洞穴环境相适应的特化或退化特征, 不具有完全的系统学意义。因此金线鲃属的系统发育关系及洞穴物种的起源需要用其他类型的数据来研究。分子数据可能会更好地解决有争议类群的系统发育关系, 因为许多分子数据相对于形态数据而言受适应性进化的影响小, 如中性位点。Xiao 等人<sup>[9]</sup>首次利用线粒体细胞色素 *b* 和 ND4 基因研究了金线鲃属的系统发育关系, 得出了与形态学工作<sup>[7,8]</sup>迥异的结果: 金线鲃属的洞穴物种并不是一个单系群, 而是分散在 5 个分支中。Xiao 等人的工作极大地增进了我们对金线鲃属系统发育的了解。但是他们选择的唯一外类群是不合适的, 因为四须鲃属不是一个单系群<sup>[10,11]</sup>; 同时他们认为由于快速的物种形成而导致这 5 个分支之间的相互关系没有解决<sup>[9]</sup>。由于没有标定点进行分子钟校正, 他们简单采用了其他鱼类线粒体蛋白质编码基因的

替代速率, 因而其分化时间的估算可能不准确。

本研究利用贝叶斯联合模型分析重建金线 属的系统发育关系。贝叶斯联合模型具有诸多优点: 它可以对联合数据分区; 对每个分区分别赋予一个最适合的模型; 同时也允许每个模型的参数有极大的变动。贝叶斯联合模型对联合数据的分析提供了极大的便利, 应会降低系统误差, 得到更准确的后验概率<sup>[12,13]</sup>。本研究测定了 5 种金线 属鱼类的细胞色素 *b* 和 ND4 基因序列, 主要目的是: ( ) 利用更多的外类群重建金线 属的系统发育关系; ( ) 利用非参数速率平滑法(nonparametric rate smoothing, NPRS)、罚分似然法(penalized likelihood, PL)及贝叶斯法 3 种松散分子钟方法来估算所得到的分支起源时间。

## 1 材料和方法

( ) 样品来源及外类群的选择。本研究所用 5 种金线 属鱼类共 9 个个体采自于广西和云南。由于 GenBank 没有花斑裸鲤(*Gymnocypris eckloni*)的 ND4 基因序列, 因此我们也测定了其 ND4 序列。赵亚辉在详尽的形态学工作基础上, 给出了一个新的分类检索系统<sup>[14]</sup>, 因此金线 属鱼类标本的鉴定按其分类系统。所有肌肉样品用 95%乙醇固定, 标本保存于中国科学院水生生物研究所淡水鱼类博物馆。根据鲤科鱼类的分子系统发育结果<sup>[15]</sup>, 我们选择鲤科中处于基部位置的斑马鱼(*Danio rerio*)作为远缘外类群, 还选择了广义鲤亚科的一些物种作为外类群: 即鲤族(Cyprinini)的鲤鱼(*Cyprinus carpio*)、黑鲫(*Carassius carassius*), 青海湖裸鲤(*Gymnocypris przewalskii*)、花斑裸鲤(*Gymnocypris eckloni*)、( *Barbus barbatus*)、宽头四须 (*Barbodes laticeps*)和异斑小 ( *Puntius ticto*); 野鲮族(Labeonini)的贝氏野鲮(*Labeo batesii*)<sup>[15]</sup>。

( ) PCR 扩增和 DNA 测序。采用标准的苯酚/氯仿抽提法<sup>[16]</sup>从 95%乙醇保存的肌肉样品中提取基因组 DNA。通过 PCR 扩增来获得细胞色素 *b* 和 ND4 基因的目的片段。细胞色素 *b* 基因 PCR 扩增引物为 Glu-F 和 Thr-R<sup>[17]</sup>; ND4 基因的引物序列为: ND4F (5'-AAC AAG ACC TCT GAT TTC GGC TCA-3') 和 ND4R (5'-TAG CTT CCA CTT GGA TTT GCA CC-3')。PCR 反应总体积为 50  $\mu$ L, 其中含 100 ng 基因组 DNA, 10  $\times$  缓冲液 5  $\mu$ L, 引物(10  $\mu$ mol/L)各 1  $\mu$ L, dNTP (每种浓度为 2.5  $\mu$ mol/L) 2  $\mu$ L, *Taq* 聚合酶 2.0 U, 最后补

足双蒸水至终体积。PCR 反应条件为: 94 预变性 3 min; 94 变性 30 s, 54 退火 30 s (ND4 基因为 51 ), 72 延伸 1 min (ND4 基因为 90 s), 共 35 个循环; 72 终延伸 8 min。每次反应设立不含 DNA 模板的空白对照组。扩增产物经 0.8%的琼脂糖凝胶电泳后, 用 OMEGA 凝胶回收试剂盒(OMEGA Bio-Tek)回收目的片段。ND4 全基因序列的测定采用纯化的 PCR 产物, 测序引物采用与 PCR 反应相同的引物及一条中间引物(5'-GAT GAG TAG GCA ATT AGT GA-3')。为了得到细胞色素 *b* 基因全序列, 以 pMD18-T (TaKaRa) 为连接载体, 以 Top 10 作为宿主菌, 对目的片段进行克隆。对筛选到的阳性克隆进行测序, 引物采用 M13 通用测序引物。内类群及外类群其他物种的细胞色素 *b* 和 ND4 基因序列均从 GenBank 下载。所有序列信息见表 1。

( ) 系统发育分析。采用 Clustal X<sup>[18]</sup>软件对 DNA 序列进行排序, 参数值为默认值, 并对排序结果进行手工校正。在 MEGA3.1<sup>[19]</sup>中把 DNA 序列翻译成蛋白质序列, 检查是否存在无义突变。利用 PAUP\* 4.0b10<sup>[20]</sup>计算序列碱基组成及对序列碱基组成的异质性进行卡方检验(chi-square test,  $\chi^2$ )。

贝叶斯联合模型分析采用 MrBayes 3.1.1<sup>[21]</sup>软件, 用 MODELTEST 3.7<sup>[22]</sup> 选择联合数据中每个分区的最适模型。选择标准用相对于广泛使用的等级似然率检验(hierarchical likelihood-ratio tests, hLRTs)具有一些优势的贝叶斯信息准则(Bayes information criterion, BIC)。等级似然率检验的一些缺陷有: (1) 需要对参数的增加或删除做任意选择; (2) 需要确定增加或移除参数的顺序; (3) 不能考虑模型选择的不确定性<sup>[23]</sup>, 因此本研究所有模型选择的标准都基于 BIC。由于不同基因和相同基因的不同密码子位点可能具有很不相同的进化模式, 按照 Brandley 等人<sup>[13]</sup>的方法, 我们采用了 8 种不同的分区策略, 用贝叶斯因子来选择最优的分区策略(表 2)。所有的分区策略均用大写字母 P 表示, 数字下标表示分区的数目, 具有同一数字下标但有不同字母下标表示具有同样的分区数目但分区策略不同。两个不同假设  $H_0$  和  $H_1$  的贝叶斯因子(Bayes factor)  $B_{01}$  等于  $H_0$  和  $H_1$  边际似然值之比, 但边际似然值很难直接计算, 可用调和平均数来代替<sup>[24]</sup>。我们用 MrBayes 中的 *sump* 命令得到调和平均数。因此, 贝叶斯因子可表示为两种不同分区策略之

表1 本研究所用标本及序列的详细信息

类元和标本	采集地 <sup>a)</sup>	GenBank 登录号 <sup>b)</sup>	
		细胞色素 <i>b</i>	ND4
<i>Sinocyclocheilus macrocephalus</i>		AY854683 AY854684	AY854740 AY854741
<i>Sinocyclocheilus oxycephalus</i>		AY854685	AY854742
<i>Sinocyclocheilus lunanensis</i>		AY854686	AY854743
<i>Sinocyclocheilus maitianheensis</i>		AY854710	AY854767
<i>Sinocyclocheilus anophthalmus</i>		AY854715	AY854772
<i>Sinocyclocheilus malacopterus</i>		AY854697, AY854699, AY854700	AY854754, AY854756, AY854757
<i>Sinocyclocheilus guishanensis</i>		AY854722	AY854779
<i>Sinocyclocheilus yangzongensis</i>		AY854725, AY854726	AY854782, AY854783
<i>Sinocyclocheilus qujingensis</i>		AY854719	AY854776
<i>Sinocyclocheilus rhinoceros</i>		AY854720	AY854777
<i>Sinocyclocheilus angustiporus</i>		AY854702	AY854759
<i>Sinocyclocheilus hyalinus</i>		AY854721	AY854778
<i>Sinocyclocheilus huaningensis</i>		AY854718	AY854775
<i>Sinocyclocheilus lateristriatus</i>		AY854703, AY854704, AY854706	AY854760, AY854761, AY854763
<i>Sinocyclocheilus tingi</i>		AY854701	AY854758
<i>Sinocyclocheilus grahami</i>		AY854694, AY854695, AY854696	AY854751, AY854752, AY854753
<i>Sinocyclocheilus microphthalmus</i>		AY854687, AY854689	AY854744, AY854746
<i>Sinocyclocheilus lingyunensis</i>		AY854691, AY854693	AY854748, AY854750
<i>Sinocyclocheilus anatirostris</i>		AY854708	AY854765
<i>Sinocyclocheilus tianeensis</i>		AY854717	AY854774
<i>Sinocyclocheilus furcodorsalis</i>		AY854709	AY854766
<i>Sinocyclocheilus halfibindus</i>		AY854723	AY854780
<i>Sinocyclocheilus altishoulderus</i>		AY854724	AY854781
<i>Sinocyclocheilus jii</i>		AY854727, AY854728	AY854784, AY854785
<i>Sinocyclocheilus macrolepis</i>		AY854729	AY854786
<i>Sinocyclocheilus macrophthalmus</i>		AY854733	AY854790
<i>Sinocyclocheilus jiuxuensis</i>		AY854736	AY854793
<i>Sinocyclocheilus bicornutus</i>		AY854730, AY854731	AY854787, AY854788
<i>Sinocyclocheilus cyphotergous</i>		AY854711	AY854768
<i>Sinocyclocheilus multipunctatus</i>		AY854712, AY854713	AY854769, AY854770
<i>Sinocyclocheilus longibarbatu</i>		AY854714	AY854771
<i>Sinocyclocheilus xunlensis</i>	广西环江	<b>EU366187, EU366190</b>	<b>EU366184, EU366185</b>
<i>Sinocyclocheilus yimenensis</i>	云南易门	<b>EU366191, EU366192</b>	<b>EU366179, EU366180</b>
<i>Sinocyclocheilus purpureus</i>	云南罗平	<b>EU366189, EU366194</b>	<b>EU366177, EU366178</b>
<i>Sinocyclocheilus maculatus</i>	云南宜良	<b>EU366193</b>	<b>EU366183</b>
<i>Sinocyclocheilus qiubeinsis</i>	云南嵩明	<b>EU366188, EU366195</b>	<b>EU366181, EU366182</b>
<b>外类群</b>			
<i>Cyprinus carpio</i>		X61010	X61010
<i>Carassius carassius</i>		NC_006291	NC_006291
<i>Barbus barbus</i>		NC_008654	NC_008654
<i>Puntius ticto</i>		NC_008658	NC_008658
<i>Labeo batesii</i>		AB238967	AB238967
<i>Gymnocypris przewalskii</i>		AB239595	AB239595
<i>Gymnocypris eckloni</i>		AY463522	<b>EU366186</b>
<i>Barbodes laticeps</i>		AY854738	AY854795
<i>Danio rerio</i>		AC024175	AC024175

a) 没有采集地信息的金线 属物种, 请参照文献[9]; b) 新测基因序列的 GenBank 登录号以黑体显示

间调和平均数之比值. 本研究中, 如果  $2 \ln \text{Bayes factor} = 10$ , 则认为备择假说得到很强烈的支持<sup>[25]</sup>. 初步分析发现  $\text{Temp} = 0.2$  (MrBayes 的默认值)时, 相

邻链之间交换不充分或者完全没有交换; 因此我们把参数 Temp 的值设为 0.04, 发现相邻链之间交换的频率在 10%~70% (MrBayes 的推荐值)之间, 此值被

用在所有的贝叶斯分析中。所有的贝叶斯分析均以随机树作为起始树, 4 条链(3 条热链 1 条冷链)同时运行  $5 \times 10^6$  代, 每 1000 代对系统树抽样一次, 其他参数的先验值设定为默认值。除了支长和拓扑结构外, 同一分析中不同分区其他参数都使用 unlink 命令来区别对待; 同时我们也使用“prset ratepr = variable”选项来允许分区内速率变异的存在。我们把似然值对代数作图来检验贝叶斯分析的收敛性, 用在线软件 AWTY<sup>[26]</sup>的 cumulative 命令把模型中所有的参数对运行代数作图来检测贝叶斯分析的收敛性。不同分区策略的贝叶斯分析均从不同起点独立运行两次。比较两次运行似然值的平均值, 同时也用 AWTY<sup>[26]</sup>的 compare 命令来比较两次运行每个分支的后验概率值。如果两次运行中上述参数很相似, 则用两次运行“burn-in”之后的树来生成一致树。

表2 本研究所采用的分区策略

分区策略	分区方法
P <sub>1</sub>	一个分区
P <sub>2A</sub>	联合数据密码子第一和第二位点; 联合数据密码子第三位点
P <sub>2B</sub>	细胞色素 <i>b</i> ; ND4
P <sub>3</sub>	对联合数据按三个密码子位点分区
P <sub>4A</sub>	细胞色素 <i>b</i> 作为一个分区, ND4 按三个密码子位点分区
P <sub>4B</sub>	细胞色素 <i>b</i> 按三个密码子位点分区; ND4 作为一个分区
P <sub>4C</sub>	细胞色素 <i>b</i> 密码子第一和第二位点; 细胞色素 <i>b</i> 密码子第三位点; ND4 密码子第一和第二位点; ND4 密码子第三位点
P <sub>6</sub>	对细胞色素 <i>b</i> 和 ND4 分别按三个密码子位点分区

( ) 分子钟检验及分化时间的估算。为了估算分化时间, 我们首先用 PAUP\* 4.0b10<sup>[20]</sup>进行似然率检验, 即替代速率恒定假设是否适合细胞色素 *b* 和 ND4 基因序列。将基于相同替代模型的具有分子钟约束的 ML 树与不受分子钟约束的 ML 树的似然对数值相比较; 如果二者差异不显著, 则分子钟假设不被拒绝。我们也用 BEAST v1.4.5<sup>[27]</sup>来研究替代速率在系统树中的行为, 采用不相关松散对数正态分子钟 (uncorrelated relaxed lognormal clock) 模型, 总共运行  $3 \times 10^6$  代, 每 1000 代进行抽样, 核苷酸替代模型选择 GTR + I +  $\Gamma$ 。用 Tracer v1.3<sup>[28]</sup>分析 BEAST 抽样产生的数据, burn-in 值为  $1.5 \times 10^6$ 。

迄今为止, 还没有金线 属鱼类化石或相关地

质事件的报道, 因此系统发育树不能在内类群中标定。我们采用在外类群的两个标定点: C1 标定点是黄河上游和青海湖隔离的地质时间 0.15 Ma<sup>[29]</sup>, 作为下限; C2 标定点是计算的鲤族和野鲮族的分歧时间<sup>[15]</sup> 15.96 Ma, 因为鲤族的结鱼属 (*Tor*) 未包括在本研究中, 此标定点作为上限。根据贝叶斯因子选择的最优分区策略所建系统树作为进行分子钟标定的树。我们用非参数速率平滑法<sup>[30]</sup>及罚分似然法<sup>[31]</sup>进行分子钟的估算, 这两种方法均在 r8s 中执行<sup>[32]</sup>。我们也用 Multidivtime<sup>[33,34]</sup>(可从 <http://statgen.ncsu.edu/thorne/> 获得) 进行贝叶斯松散分子钟的计算。这两个软件基于不同的原理: r8s 根据树的枝长来计算, Multidivtime 中执行的贝叶斯方法依据导入树的拓扑结构和相应的序列数据来计算。对 NPRS 和 PL 法, 分别选择 Powell 和 TN 方法来进行优化, PL 方法中的最佳平滑值由交叉验证法 (cross-validation) 获得。对贝叶斯分子钟估算而言, 我们对最优的拓扑结构进行修剪, 使每个物种只有一个单倍型。贝叶斯方法可整合不同谱系和不同基因之间进化速率的差异。把联合数据按基因分为两个分区。按照 Rutschmann<sup>[34]</sup>, Yang 和 Yoder<sup>[35]</sup>的方法, 分为 3 步: (1) 用 PAML v3.14<sup>[36]</sup>软件包中的 Baseml 程序估算  $F84 + \Gamma$  模型中的相关参数: 碱基频率、转换/颠换比值  $\kappa$ ,  $\Gamma$  分布的形状参数  $\alpha$ ; (2) 用 Multidivtime 软件包中的 Estbranches 程序估算系统树的支长及其支长的方差-协方差矩阵, 以 Baseml 的输出结果作为 Estbranches 的输入; (3) 把作为所有物种外类群的斑马鱼从系统树中去掉, Estbranches 的输出结果作为 Multidivtime 的输入, 马尔科夫链的蒙特卡洛方法 (Markov chain Monte Carlo process, MCMC) 设置为  $2 \times 10^6$  代, 每 100 代抽样一次, burn-in 设定为  $2 \times 10^5$  代。为确保 MCMC 算法的收敛性, MCMC 从不同起点独立运行两次, 得到节点分化时间的标准偏差 (standard deviations, SD) 及 95% 置信区间 (credibility intervals, CI)。在运行 MCMC 之前, 我们先对 root age, root rate 和 rate autocorrelation 的平均值及标准偏差等参数设定如下: 以 18 Ma (SD = 1.8 Ma) 作为 tree root 到 tip 分化时间的期望值 (鲤亚科的出现时间<sup>[15]</sup>); 每位点每百万年 0.007 的替换速率 (SD = 0.0007) 作为 tree root 节点的进化速率; 分支自主相关程度的参数 ( $\nu$ ) 设置为 0.08 (SD = 0.08)。我们按照 Thorne 和 Kishino<sup>[33]</sup>的建议调整了  $\nu$  的值, 使  $\nu$  与 root age 之积在 1 与 2 之间。

## 2 结果

### 2.1 细胞色素 *b* 和 ND4 基因的序列特征

我们测定了 9 个个体的细胞色素 *b* 全基因序列, 共得到 8 个单倍型. 与其他从 GenBank 下载的细胞色素 *b* 全基因序列比对后得到 1140 个位点, 包含 537 个可变位点, 其中 434 个位点为简约信息位点, 比例分别为 47.1% 和 38.1%. A, T, C, G 的平均含量分别为 29.5%, 27.9%, 27.9%, 14.7%, 呈反 G 偏倚. 异质性检验表明密码子第 1 和 2 位点是同质的, 但第 3 位点具有异质性: 第 1 位点,  $\chi^2=34.036$ ,  $df=180$ ,  $P=1.000$ ; 第 2 位点,  $\chi^2=3.756$ ,  $df=180$ ,  $P=1.000$ ; 第 3 位点,  $\chi^2=316.585$ ,  $df=180$ ,  $P<0.001$ . 但所有位点的组成是同质的:  $\chi^2=130.112$ ,  $df=180$ ,  $P=0.998$ .

测定了金线属 9 个个体及花斑裸鲤 1 个个体的 ND4 全基因序列(1381 bp), 得到 10 个单倍型, 以 T 作为不完全终止密码子. 与下载的 ND4 基因序列比对后得到 1032 个位点, 包含 484 个可变位点, 其中 397 个位点为简约信息位点, 比例分别为 46.9% 和 38.5%. A, T, C, G 的平均含量分别为 31.0%, 27.3%, 27.1%, 14.6%, 呈反 G 偏倚. 异质性检验表明密码子 3 个位点的组成均是同质的: 第 1 位点,  $\chi^2=25.028$ ,  $df=180$ ,  $P=1.000$ ; 第 2 位点,  $\chi^2=3.979$ ,  $df=180$ ,  $P=1.000$ ; 第 3 位点,  $\chi^2=167.505$ ,  $df=180$ ,  $P=0.739$ . 细胞色素 *b* 及 ND4 的同质性表明我们所用的数据不会对系统树的构建产生负面影响.

### 2.2 联合数据的分区及金线属的系统发育

联合数据共包含 2172 个碱基. 不同分区的最适模型见表 3. 由于 MrBayes 中没有 TrN 和 TrNef 模型, 因此我们用更通用的 GTR 模型来代替.

所有的贝叶斯联合模型分析的似然值在  $2 \times 10^5$  代就达到了平衡; 但用 AWTY<sup>[26]</sup>的 *cumulative* 命令把模型中的所有参数对运行代数作图表明 MCMC 在  $2.5 \times 10^6$  代才达到平衡(未显示). 每个分区策略两次运行的似然值都很相似, 其他参数也没有显著性差异. 因此丢弃每次运行的初始 2500 棵树, 用剩余的 5002 棵树生成一致树.

各个分区策略两两比较之贝叶斯因子及似然值平均数见表 4. 从表中可以看出, 对联合数据按 3 个密码子位点分区的策略(P<sub>3</sub>)是最好的. 由于每个贝叶斯联合模型分析所得到的系统树拓扑结构几乎完全一致, 因此我们以 P<sub>3</sub> 生成的贝叶斯树(图 1)来讨论金

表 3 不同分区及其最适进化模型和所包含位点数

分区	进化模型	分区所包含位点数
ND4	TrN + I + G	1032
ND4 第 1 位点	TrNef + G	344
ND4 第 2 位点	HKY + I + G	344
ND4 第 3 位点	TrN + I + G	344
ND4 第 1 和第 2 位点	HKY + I + G	688
cyt <i>b</i>	TrN + I + G	1140
cyt <i>b</i> 第 1 位点	K80 + G	380
cyt <i>b</i> 第 2 位点	HKY + I	380
cyt <i>b</i> 第 3 位点	GTR + I + G	380
cyt <i>b</i> 第 1 和第 2 位点	HKY + I + G	760
联合数据	TrN + I + G	2172
联合数据第 1 位点	K80 + I + G	724
联合数据第 2 位点	HKY + I + G	724
联合数据第 3 位点	GTR + I + G	724
联合数据第 1 和第 2 位点	HKY + I + G	1448

线属的系统发育关系. 金线属的单系性得到强烈的支持, 其后验概率为  $PP=0.96$ ; 金线属可分为 6 个分支. 季氏金线 (*S. jii*) 在金线属中处于基部位置; 其他金线属鱼类形成一强烈支持的单系群 ( $PP=1.00$ ), 为 *S. jii* 的姐妹群. 除 *S. jii* 外的其他金线属鱼类可分为 5 个分支: 分支 A, C, D 和 E 均得到强烈的支持, 后验概率均为  $PP=1.00$ ; 分支 B 的后验概率不高, 为 0.74. 分支 D 和 E 的所有物种均为典型洞穴鱼类, 后者形成分支 A, B, C 和 D 的姐妹群 ( $PP=1.00$ ). 分支 A 所包含的物种是最多的.

### 2.3 分子钟的估算

对数似然比检验表明, 细胞色素 *b* ( $\chi^2=152.33558$ ,  $df=59$ ,  $P<<0.001$ ) 和 ND4 ( $\chi^2=102.00113$ ,  $df=59$ ,  $P<0.001$ ) 的替代速率恒定假设均不成立. 通过 Tracer v1.3 分析得到参数 uclid 的标准偏差为 0.519, 速率变异系数为 0.512, 均小于 1, 表明联合数据具有中等的速率异质性. PL 法中, 通过交叉验证法得到的最优平滑值为 0.0001. 贝叶斯联合模型 P<sub>3</sub> 生成的系统树作为进行分子钟标定的树. 用 Multidivtime 得到的超量树见图 2, 表 5 列出了主要节点的分化时间, 3 种方法所计算的结果很一致. 金线属的起源时间为 11 Ma 左右(95% CI 为 13.23~8.00 Ma), 节点 2~5 随后快速出现, 表示在这一段时间有快速的物种形成事件. 除了分支 C 从中上新世到晚上新世开始辐射, 其他分支都从晚中新世开始辐射.

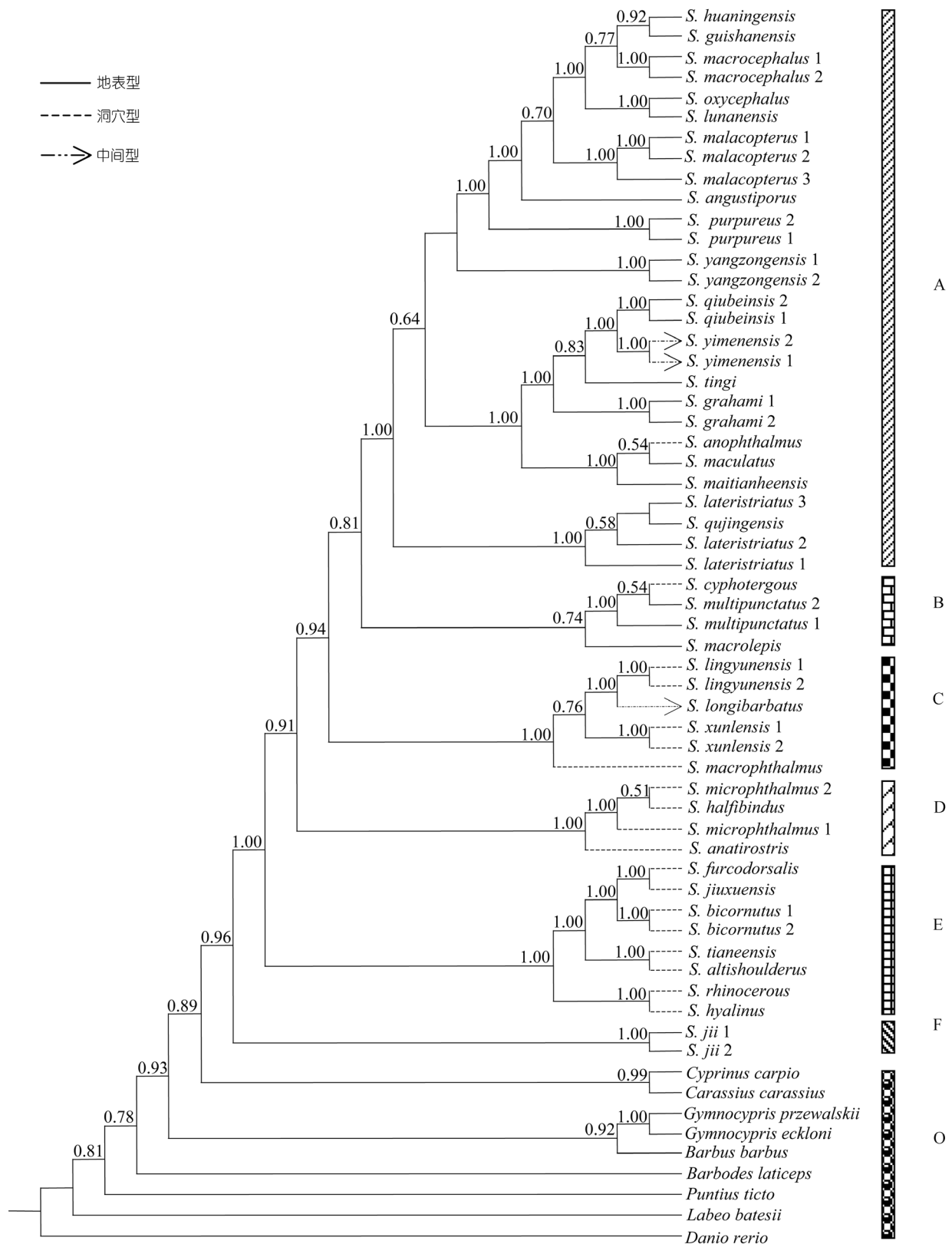


图1 通过贝叶斯联合模型(P<sub>3</sub>)构建的50%一致树  
 节点处的数字代表后验概率(小于50%的未显示)

表4 各个分区策略之贝叶斯因子及似然值平均数

分区策略	P <sub>1</sub>	P <sub>2A</sub>	P <sub>2B</sub>	P <sub>3</sub>	P <sub>4A</sub>	P <sub>4B</sub>	P <sub>4C</sub>	P <sub>6</sub>
-lnL	20445.47	19555.93	20437.85	19458.26	20024.37	19898.88	19542.06	19477.12
P <sub>1</sub>	-							
P <sub>2A</sub>	1761.96	-						
P <sub>2B</sub>	-12.62	-1774.58	-					
P <sub>3</sub>	1966.94	204.98	1979.56	-				
P <sub>4A</sub>	822.28	-939.68	834.9	-1144.66	-			
P <sub>4B</sub>	1080.22	-681.74	1092.84	-886.72	257.94	-		
P <sub>4C</sub>	1788.18	26.22	1800.8	-178.76	965.9	707.96	-	
P <sub>6</sub>	1891.90	129.94	1904.52	-75.04	1069.62	811.68	103.72	-

a) 左下角数字为贝叶斯因子

### 3 讨论

#### 3.1 贝叶斯联合模型分析的性能及分区的选择

从表4可以看出, 贝叶斯联合模型分析的确极大地提高了似然值. 但随着分区数目的增加, 似然值并不一定得到提高; 每个分区如何划分更重要, 这支持了 Brandley 等人<sup>[13]</sup>、Guo 和 Wang<sup>[37]</sup>的结论. 例如, Xiao 等人<sup>[9]</sup>对联合数据采用的按基因分区的 P<sub>2B</sub>, 并不优于 P<sub>1</sub>; P<sub>4A</sub>, P<sub>4B</sub>, P<sub>4C</sub> 和 P<sub>6</sub> 虽然分区数目比 P<sub>3</sub> 多, 但 P<sub>3</sub> 的似然值是最高的. 就基因而言, 对细胞色素 *b* 按密码子位点分区(P<sub>4B</sub>)比对 ND4 按相应位点分区(P<sub>4A</sub>)对似然值提高的贡献大一些, 前者与后者相比提高了 125.49 个对数似然单位. 对联合数据按密码子位点分区的 P<sub>2A</sub>, P<sub>3</sub>, P<sub>4C</sub> 和 P<sub>6</sub>, 比按照基因分区的策略 P<sub>2B</sub>, P<sub>4A</sub> 和 P<sub>4B</sub> 对似然值提高的贡献大得多, 表明线粒体同一条链上不同蛋白质基因相同密码子位点具有相似的进化模式. 贝叶斯因子表明, P<sub>3</sub> 比其他任何一种分区策略提高的似然值都大. 在 2lnBayes factor 10 的标准下, 所有的分区策略都是互不相同的, 但是拓扑结构都很相似, 只在某些节点的后验概率上有差异, 尤其是一些深部节点. 例如, P<sub>2B</sub> 的深部节点的后验概率 0.85, 0.96, 0.51 和 0.51 分别被 P<sub>3</sub> 相应节点的后验概率 0.91, 0.94, 0.81 和 0.64 代替. 正如 Brandley 等人<sup>[13]</sup>、Guo 和 Wang<sup>[37]</sup>指出的那样, 我们也认为现行的 2lnBayes factor 10 标准是否对额外的分区给予了足够的罚金.

#### 3.2 金线鲃属的系统发育

我们的系统发育结果与基于形态工作的系统发育假设不相同<sup>[7,8]</sup>, 但与 Xiao 等人<sup>[9]</sup>的结果部分一致: 洞穴物种是一个多系群, 散布在 5 个分支中, 不支持王大忠等人<sup>[8]</sup>提出的洞穴物种是单系群的假设; 也不支持单乡红和乐佩琦提出的洞穴物种是二分支的假

设<sup>[7]</sup>, 这说明有些形态性状不具有完全的系统学意义. 由于鲤鱼和黑鲫与金线 属的关系更近, 因此宽头四须 并不是一个最合适的外类群. 我们得到的分支 D 和 E 分别与 Xiao 等人<sup>[9]</sup>得到的分支 和 相同; 分支 C 由 *S. xunlensis* 与 Xiao 等人<sup>[9]</sup>得到的分支 形成; 本研究新增另外 4 个物种与 Xiao 等人<sup>[9]</sup>得到的分支 形成分支 A. 分支 B 是本研究与 Xiao 等人的一个主要不同之处: Xiao 等人的分支 包括 *S. multipunctatus* 和 *S. cyphotergous*, 在本研究中, 分支 B 包括这两个物种和 *S. macrolepis*. 虽然分支 B 的后验概率不高; 但在我们的 MP 树和 ML 树中, 分支 B 的支持率均大于 50% (未发表数据), 说明分支 B 是可靠的. Xiao 等人没有解决 *S. macrolepis* 的位置, 可能是因为他们选择了不合适的外类群.

Xiao 等人与本研究的另一个不同之处在于这 6 个分支之间的相互关系. 在 MP 树中, 其相互关系为 (((((A,C),B),D),E),F), 并且根据 AU, KH 和 SH 检验, MP 树与贝叶斯树((((A,B),C),D),E),F)并没有显著差别(未发表数据). 因此, 只有分支 A, B 和 C 的相互关系没有解决. 相比之下, Xiao 等人<sup>[9]</sup>的 5 个分支之间的相互关系都没有解决. 因此, 我们认为 Xiao 等人没有解决分支 和 在系统树中的位置不能归咎于金线 属的快速分化<sup>[9]</sup>.

*S. lunanensis* 和 *S. halfibindus* 的物种有效性还具有争议<sup>[7,9,14]</sup>. 本研究中, *S. halfibindus* 与 *S. microphthalmus* 两个单倍型之间的细胞色素 *b* 基因 *p* 距离分别为 0.526%和 0.702%, ND4 基因的 *p* 距离分别为 0.388%和 0.484%; *S. lunanensis* 和 *S. oxycephalus* 细胞色素 *b* 和 ND4 基因 *p* 距离分别为 0.175%和 0.097%. 由于它们的遗传距离非常小, 形态上又非常相似, 因此我们支持 *S. halfibindus* 是 *S. microphthalmus* 的同物异名; *S. lunanensis* 是 *S. oxycephalus* 的同物异名<sup>[7,14]</sup>.

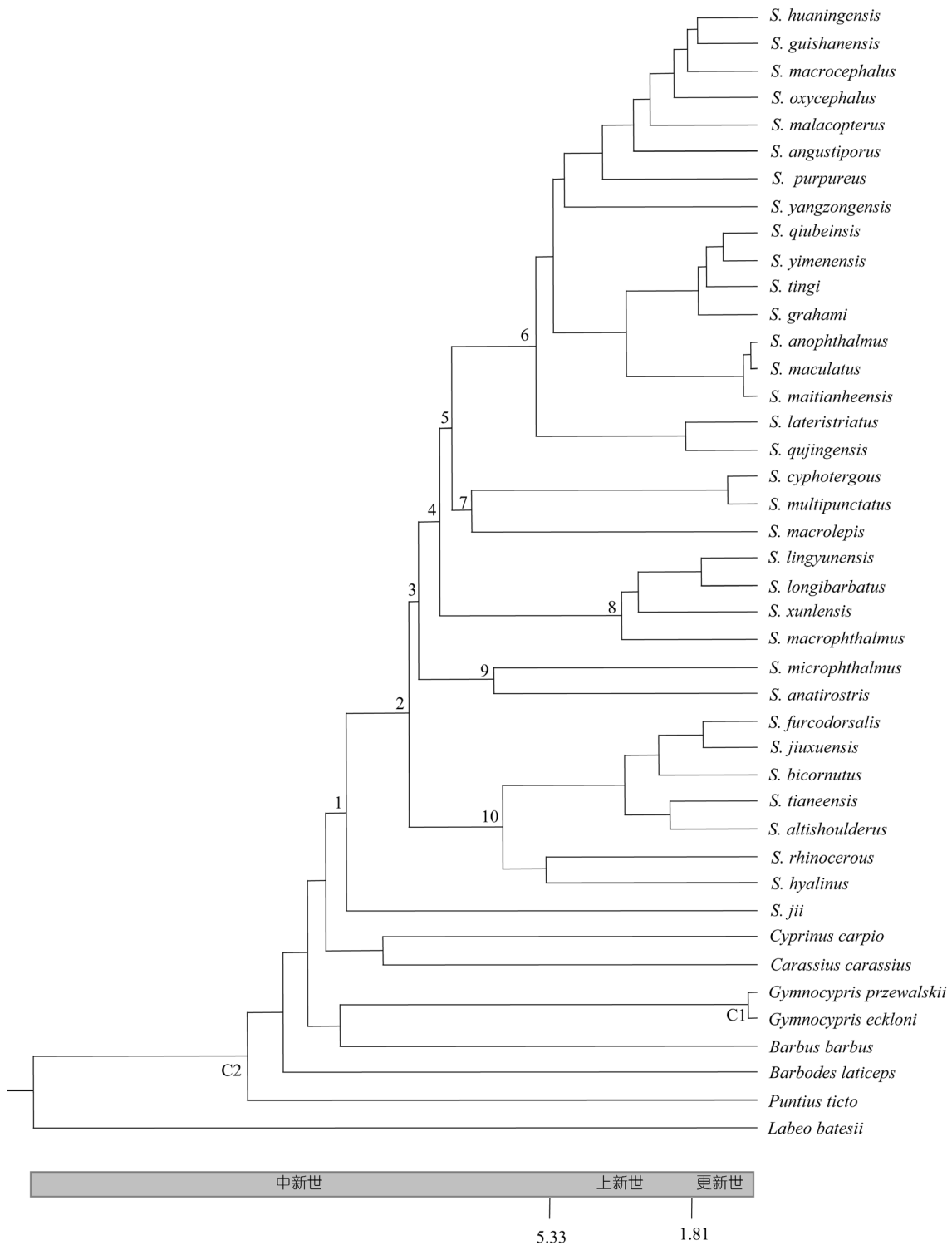


图 2 从 Multidivtime 得到的超量树

支长代表时间, C1 和 C2 代表标定点

### 3.3 金线鲃属鱼类的起源及分化

与 Xiao 等人的研究结果相比, 本研究估算的金

线 属起源时间及其内部分支起源时间均比其估算的要久远. 例如, 分支 D 和 E 估算的分歧时间其平均



表5 3种方法计算的主要节点的分化时间

节点	r8s		Multidivtime	
	NPRS	PL	均值(SD)	95% CI
1	11.51	11.41	10.61(1.35)	(8.00, 13.23)
2	9.36	9.16	8.99(1.34)	(6.42, 11.64)
3	8.79	8.58	8.73(1.33)	(6.19, 11.37)
4	8.09	7.87	8.20(1.32)	(5.68, 10.80)
5	7.68	7.45	7.88(1.32)	(5.37, 10.51)
6	5.63	5.27	5.71(1.20)	(3.50, 8.12)
7	7.12	6.90	7.38(1.31)	(4.89, 10.01)
8	2.97	2.85	3.49(1.08)	(1.69, 5.88)
9	6.29	6.10	6.80(1.36)	(4.26, 9.54)
10	6.44	5.89	6.54(1.30)	(4.15, 9.22)

值分别为 6.80 和 6.54 Ma, 95%置信区间分别为 4.26~9.54, 4.15~9.22 Ma; 而 Xiao 等人得到的分支和 其分歧时间 95% HPD (highest posterior density) 分别为 2.254~8.177 和 1.94~6.299 Ma. 我们估算的分化时间与王大忠和陈宜瑜<sup>[38]</sup>的假设部分一致, 即金线属鱼类的原始祖先在第三纪晚期已存在于云贵高原一带, 在第三纪结束时已发生了一定程度的物种分化. 但同时我们的研究结果也表明本属鱼类的起源及分化时间比王大忠和陈宜瑜估计的要久远得多.

金线属的最近共同祖先(most recent common ancestor, MRCA)出现在 11 Ma 左右, 随后在 9~8 Ma 发生了 4 次快速的分支形成事件(图 2, 节点 2, 3, 4 和 5). 我们估算的分歧时间可为青藏高原隆起过程的假说提供佐证, 因为金线属鱼类扩散能力弱, 分布局

限, 云贵高原的隆升与青藏高原的隆升密切相关. 青藏高原隆起的假说有两种: 一种认为在 20 Ma 左右青藏高原开始了快速的隆起及夷平, 直到 10~8 Ma 才开始显著的隆起从而达到有意义的高度<sup>[39]</sup>; 另一种观点则认为青藏高原在 8 Ma 前就已经达到了最大高度, 随后由于张裂(extensional faulting)而降低, 在 3.6 Ma 左右开始了近期的快速隆升<sup>[40]</sup>. 按照前一种假说, 东亚季风肇始于 9~8 Ma, 夏季风会给东亚地区带来丰沛的降水<sup>[41]</sup>. 因此, 本研究支持前一种假说.

众所周知, 大气中二氧化碳含量的变化对全球气温具有巨大而可以预测的变化. 有趣的是, 我们发现这一时期全球二氧化碳含量的变化与金线属鱼类的分化具有很大的—致性. 自从早中新世以来(大约 24 Ma 左右), 大气中二氧化碳的含量比以前地质时期二氧化碳的含量要低——保持在 500  $\mu\text{L/L}$  以下, 并且相对而言也比以前更加稳定<sup>[42]</sup>; 但是二氧化碳含量的波动仍是相当可观的. 研究表明, 从 14 Ma 到 9 Ma, 大气中二氧化碳含量处于稳步升高的过程, 随后一直保持工业化前的水平<sup>[42,43]</sup>, 均大于 210  $\mu\text{L/L}$ <sup>[42]</sup>. 这 10 个节点的估算的分化时间与大气中二氧化碳含量的波动是大体相关的. 在北半球的大冰期来临之前, 亚洲大陆温暖而湿润, 加上高浓度的二氧化碳, 岩石的风化作用可能因此而加强了. 因此, 金线属的分化与这一时期的环境变化可能紧密关联, 但是这需要收集更多的证据.

致谢 李再云、陈自明、陶进能在采样过程中予以帮助; 李维贤提供部分标本; 郭宪光和 Brandley 在贝叶斯联合模型分析中给予热心帮助; 王绪祯博士对论文写作提出了宝贵意见, 在此一并表示感谢.

## 参考文献

- Culver D C, Kane T C, Fong D W. Adaptation and Natural Selection in Caves. Cambridge: Harvard University Press, 1995
- Romero A. The Biology of Hypogean Fishes. Dordrecht: Kluwer Academic Publishers, Netherlands, 2001
- Protas M E, Hersey C, Kochanek D, et al. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet, 2006, 38(1): 107—111
- Jeffery W R. Adaptive evolution of eye degeneration in the Mexican blind cavefish. J Hered, 2005, 96(3): 185—196
- Dowling T E, Martasian D P, Jeffery W R. Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. Mol Biol Evol, 2002, 19(4): 446—455
- Zhao Y H, Watanabe K, Zhang C G. *Sinocyclocheilus donglanensis*, a new cavefish (Teleostei: Cypriniformes) from Guangxi, China. Ichthyol Res, 2006, 53: 121—128
- 单乡红, 乐佩琦. 金线鲃属鱼类系统发育研究(鲤形目: 鲤科: 鲃亚科). 动物学研究, 1994, 15(增刊): 36—44
- 王大忠, 陈宜瑜, 李学英. 金线鲃属的系统发育分析(鲤形目: 鲤科: 鲃亚科). 遵义医学院学报, 1999, 22: 1—6
- Xiao H, Chen S Y, Liu Z M, et al. Molecular phylogeny of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) inferred from mitochondrial

- DNA sequences. *Mol Phylogenet Evol*, 2005, 36(1): 67—77
- 10 Chen X Y, Yang J X. A systematic revision of “*Barbodes*” fishes in China. *Zool Res*, 2003, 24: 377—386
- 11 Kottelat M. Nomenclature of the genera *Barbodes*, *Cyclocheilichthys*, *Rasbora* and *Chonerhinos* (Teleostei: Cyprinidae and Tetraodontidae), with comments on the definition of the first reviser. *Raffles Bull Zool*, 1999, 47: 591—600
- 12 Nylander J A A, Ronquist F, Huelsenbeck J P, et al. Bayesian phylogenetic analysis of combined data. *Syst Biol*, 2004, 53: 47—67
- 13 Brandley M C, Schmitz A, Reeder T W. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol*, 2005, 54: 373—390
- 14 赵亚辉. 中国特有洞穴金线鲃属(鲤形目: 鲤科)鱼类的研究——物种多样性、系统演化和动物地理. 博士学位论文. 北京: 中国科学院研究生院, 2006
- 15 Wang X Z, Li J B, He S P. Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. *Mol Phylogenet Evol*, 2007, 42: 157—170
- 16 Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989
- 17 Zardoya R, Doadrio I. Phylogenetic relationships of Iberian cyprinids: Systematic and biogeographical implications. *Proc Biol Sci*, 1998, 265: 1365—1372
- 18 Thompson J D, Gibson T J, Plewniak F, et al. The Clustal X windows interface: Flexible strategies for multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25: 4876—4882
- 19 Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*, 2004, 5: 150—163
- 20 Swofford D L. PAUP\*: Phylogenetic analysis using Parsimony (\*and other methods), Version 4. Sunderland, Sinauer Associates, 2002
- 21 Ronquist F, Huelsenbeck J P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 2003, 19: 1572—1574
- 22 Posada D, Crandall K A. Modeltest: Testing the model of DNA substitution. *Bioinformatics*, 1998, 14: 817—818
- 23 Posada D, Buckley T R. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol*, 2004, 53: 793—808
- 24 Newton M A, Raftery A E. Approximate Bayesian inference with the weighted likelihood bootstrap. *J R Stat Stoc*, 1994, 56: 3—48
- 25 Kass R, Raftery A. Bayes factor. *J Am Stat Assoc*, 1995, 90: 773—795
- 26 Wilgenbusch J C, Warren D L, Swofford D L. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available at <http://ceb.csit.fsu.edu/awty>, 2004
- 27 Drummond A J, Rambaut A. BEAST v1.4.5. University of Oxford, Oxford. Available at <http://evolve.zoo.ox.ac.uk/beast>, 2005
- 28 Rambaut A, Drummond A J. Tracer. Version 1.3, Available at <http://evolve.zoo.ox.ac.uk>, 2003
- 29 李吉均, 方小敏, 潘保田, 等. 新生代晚期青藏高原强烈隆起及其对周边环境的影响. *第四纪研究*, 2001, 21: 381—391
- 30 Sanderson M J. A nonparametric approach to estimating divergence times in the absence of rate constancy. *J Mol Evol*, 1997, 44: 1218—1231
- 31 Sanderson M J. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol Biol Evol*, 2002, 19: 101—109
- 32 Sanderson M J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 2003, 19: 301—302
- 33 Thorne J L, Kishino H. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol*, 2002, 51: 689—702
- 34 Rutschmann F. Bayesian molecular dating using PAML/multidivtime. A step-by-step manual. University of Zurich, Switzerland. Available at <http://www.plant.ch>, 2005
- 35 Yang Z, Yoder A D. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst Biol*, 2003, 52: 705—716
- 36 Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comp Appl Biosci*, 1997, 13: 555—556
- 37 Guo X G, Wang Y Z. Partitioned Bayesian analyses, dispersal-vicariance analysis, and the biogeography of Chinese toad-headed lizards (Agamidae: Phrynocephalus): A re-evaluation. *Mol Phylogenet Evol*, 2007, 45: 643—662
- 38 王大忠, 陈宜瑜. 金线鲃属鱼类的起源及其适应演化. *水生生物学报*, 2000, 24: 630—634
- 39 Harrison T M, Copeland P, Kidd W S F, et al. Raising Tibet. *Science*, 1992, 255: 1663—1670
- 40 李吉均, 方小敏, 马海洲, 等. 晚新生代黄河上游地貌演化与青藏高原隆起. *中国科学 D 辑: 地球科学*, 1996, 26: 316—322
- 41 An Z, Kutzbach J E, Prell W L, et al. Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. *Nature*, 2000, 411: 62—66
- 42 Pearson P N, Palmer M R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature*, 2000, 406: 695—699
- 43 Pagani M, Freeman K H, Arthur M A. Late Miocene atmospheric CO<sub>2</sub> concentrations and the expansion of C<sub>4</sub> grasses. *Science*, 1999, 285: 876—879