论文

基于分区贝叶斯法和最大似然法的鲤科鱼类系统发育分析及其系统学意义

王绪祯^①, 甘小妮^①, 李俊兵^①, MAYDEN Richard L.^②, 何舜平^{①*}

① 中国科学院水生生物多样性与保护重点实验室, 中国科学院水生生物研究所, 武汉 430072;

② Department of Biology, Saint Louis University, Saint Louis, Missouri 63103-2010, USA

* 联系人: E-mail: clad@ihb.ac.cn

收稿日期: 2012-04-25; 接受日期: 2012-07-16

国家自然科学基金(批准号: 30770300)、中国科学院知识创新工程(批准号: KSCX2-EW-Q-12)和美国自然科学基金(批准号: EF-0431326)资助项目

摘要 鲤科是淡水鱼类中最大的科, 但其高阶元类群之间的系统发育关系目前尚不明确. 本研究选择核基因组中的重组活化基因 2、线粒体基因组中的 *16S rRNA* 基因和细胞色素 *b* 基因作为遗传标记, 重建鲤科鱼类的系统发育关系. 研究结果显示, 分区策略对解决鲤科鱼类的系统发育关系和提升其节点支持率并没有显著效果, 因此在所有分析中, 不分区策略是最佳选择. 贝叶斯分析结果显示, 鲤科中单系类群间关系为: (鲤亚科, 野鲮亚科), ((鲮亚科, ((雅罗鱼亚科, 丁鲮亚科), 鮡亚科)), 鲃亚科). 分化时间估算表明, 鲤科鱼类的主要类群的分化集中在渐新世晚期到中新世晚期. 本研究中的系统发育分析结果有助于促进对鲤科鱼类进化历史的认识.

关键词

鲤科
系统发育
分子时间估算
分区分析

鲤科(Cyprinidae)是淡水鱼类中最大的科, 约 220 属 2420 种^[1]. 鲤科鱼类物种数目众多, 分布范围广, 外部形态差异非常显著, 这都使得鲤科鱼类的分类学研究比较困难^[2], 对其进行分支分析也是一项具有挑战性的工作. Hensel^[3]已经就鲤科鱼类的分类学历史做过了详细描述. 基于形态学或解剖学性状, 分类学家们识别出了数目不同的亚科^[2,4-8], 但对亚科间的系统关系缺乏了解. 因为这些亚科的定义通常比较模糊, 或者只有少数形态学性状支持亚科的划分^[2].

鲤科鱼类传统上被划分为 2 个系群, 即鲤(鲃)系(Cyprinine lineage)和雅罗鱼系(Leuciscine lineage).

然而, 形态学研究在鲤科亚科的数目及其单系性等关键问题上并未取得一致, 因此对人们了解鲤科鱼类科下阶元间关系的帮助十分有限. 继陈湘舜等人^[8]报道了鲤科亚科间关系的分支系统学研究结果后, Cavender 和 Coburn^[9]以及 Howes^[12]也对鲤科鱼类进行过分类, 而且都包括了单型属——丁鲮属(*Tinca*). 就丁鲮亚科(Tincinae)、波鱼亚科(Rasborinae)和鮡亚科(Gobioninae)的系统位置而言, 这些研究提出了相互矛盾的观点.

随着分子系统发育分析在鲤科鱼类研究中广泛运用, 大多数欧洲鲤科鱼类的分子系统学研究取得了

英文版见: Wang X Z, Gan X N, Li J B, et al. Cyprinid phylogeny based on bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. *Sci China Life Sci*, 2012, 55, in press

一致结果^[10-16]. 相关研究都否定了雅罗鱼亚科的单系性, 而建议将 *Alburninae* 置于其中; 不支持 Howes 的鲤科二分法^[2], 即有胡须的鲤类(包括鲤亚科(*Cyprininae*)、鲃亚科、波鱼亚科)和胡须缺失或零星分布的雅罗鱼类(包括鲮亚科(*Acheilognathinae*)、雅罗鱼亚科(*Leuciscinae*)、鲃亚科(*Cultrinae*)、*Alburninae* 亚科). 由于亚洲水体中鲤科鱼类的物种多样性最为丰富^[17], 因此包括了亚洲鲤科鱼类的系统发育分析将会促进鲤科系统学研究取得进展^[18-20]. 基于细胞色素 *b* (cytochrome *b*, *Cytb*) 基因序列分析, Cunha 等人^[18] 在鲤科中识别出了一个亚洲类群, 包括鲃类(*cultrins*)、鲮类(*acheilognathins*)、鲃类(*gobionins*)和鲮类(*xenocyprins*). 其他东亚鲤科鱼类分子系统学研究表明, 鲤科包含了 2 个主要系群, 而且东亚鲮-鲃类群及其相关类群共同构成单系^[19,20]. 以上分子系统发育研究都是基于线粒体部分 DNA 序列, 因此得到的分支树在解决鲤科系统发育关系和辨别不同系统发育假设中发挥的作用非常有限.

鲤科的分类系统基本上基于形态学研究, 因而缺乏一个完整的系统发育框架. 由于取样上的制约, 近来的分子系统学研究对鲤科分类系统所做的修订也十分有限. 鲤科系统发育关系和分类学中的一些关键问题仍需进一步澄清. 首先, 鲤科中大多数已命名亚科的单系性并没有得到分子或形态证据的证实. 对于形态学研究识别的一些亚科, 如波鱼亚科^[8]和雅罗鱼亚科^[2,4], 它们的单系性没有得到分子证据的支持^[18-20]. 其次, 以前的研究在鲤科中波鱼亚科、丁鲷亚科和鲮亚科的系统位置问题上没能取得一致. 而近来的分子分析中, 因为相应节点的统计检验支持率偏低, 鲤科亚科间的关系也没有得以解决. 再次, 尽管鲤科的二分支分类, 即雅罗鱼系和鲤系(或鲃系)已广为接受, 但这两个分支内部的高阶元类群间的关系还需要进一步研究.

东亚鲤科分子系统发育分析结果与传统形态学结果有着非常明显的差异, 特别是在鲤科的亚科划分问题上^[19,20]. 因此, 系统发育研究中对亚洲鲤科鱼类进行广泛取样, 将有助于解决鲤科的系统发育分类学. 除了广泛的样品采集之外, 本研究还采用了重组活化基因 2 (recombination activating gene 2, *RAG2*)、线粒体 *16S rRNA* 和 *Cytb* 基因的联合数据来重建鲤科系统发育关系.

为了分析由多个基因组成的 DNA 序列数据, 分

区系统发育分析方法近年来非常流行. 分区系统发育分析能够针对联合数据中的不同子集采用不同的核苷酸取代模型(及相应的参数), 从而有助于探索各个分区特异的进化模型, 减少系统误差, 以提高系统发育推断的精确性. 分区系统发育分析方法以前只是基于贝叶斯(Bayes)原理^[21], 最近又发展出了基于最大似然法(maximum likelihood, ML)的分区分析方法^[22]. 此外, 一个合理的分区策略通常要求对联合数据集要采用合理的模型, 而不是过度进行分区. 因为过度分区(或过多的参数体系)可能会导致参数不能识别、变量增加、后验值的不正常分布以及先验值的过度影响^[23]. 目前, 除了贝叶斯因子法之外, 还有一些标准用于分区贝叶斯分析中分区策略的选择, 这些标准在分析中施加了明确的参数化惩罚政策^[23].

本研究对分区数据分别采用贝叶斯和 ML 分析法来重建鲤科的系统发育关系, 并利用松散分子钟方法来估算鲤科中物种分化事件的发生时间. 本研究的目的如下: (1)展示联合数据的分区系统发育分析如何影响鲤科鱼类系统发育关系的重建; (2)检验目前分类学上识别的鲤科不同亚科的单系性; (3)探讨识别的单系类群在鲤科分类学中的意义.

1 材料与方法

1.1 样品取样与 DNA 抽提

本研究样品包括 103 种鲤科鱼类, 涵盖了鲤科鱼类全部的主要形态类群和 12 个亚科^[4]. 基于鲤形目(*Cypriniformes*)是单系类这一普遍认同观点^[24,25], 本研究选择了除鲤科之外的 6 种鲤形目鱼类作为外类群, 它们分别来自亚口鱼科(*Catostomidae*)、平鳍鳅科(*Balitoridae*)、鳅科(*Cobitidae*)和双孔鱼科(*Gyrinocheilidae*) (表 1).

野外采集的肌肉或鳍条组织用 95% 乙醇固定, -20℃ 条件下保存. 采用标准的酚-氯仿法来抽提鲤科鱼类的总基因组 DNA^[26].

1.2 DNA 序列收集与排序

核基因和线粒体基因的 PCR 扩增采用已经发表的引物, 并对其中的部分引物做了优化^[27-29]. 50 μL 的混合反应体系包括: 100 ng DNA 模板, 5 μL 10× 反应缓冲液, 2 μL dNTPs (各 2.5 mmol/L), 2.0 U DNA 聚合酶, 正反向引物各 1 μL (引物浓度为 10 μmol/L).

表 1 本研究所用的分类单元及其序列数据的 GenBank 登录号^{a)}

亚科	分类单元	采集地	GenBank 登录号		
			RAG2	16S rRNA	Cytb
鲃亚科	<i>Acrossocheilus bejiangensis</i>	融安, 广西	DQ366967	DQ845869	—
	<i>Acrossocheilus elongates</i>	融安, 广西	DQ366979	GQ406254	—
	<i>Acrossocheilus hemispinus</i>	融安, 广西	DQ366986	DQ845867	GQ406312
	<i>Balantiocheilus melanopterus</i>	水族馆	DQ366933	GQ406255	—
	<i>Barbodes huangchuchieni</i>	勐腊, 云南	DQ366952	GQ406256	—
	<i>Barbodes vernayi</i>	勐腊, 云南	DQ366987	DQ845870	GQ406313
	<i>Barbonymus schwanenfeldii</i>	水族馆	DQ366961	DQ845906	AF180823*
	<i>Barbus barbus</i>	法国	DQ366990	DQ845879	AB238965*
	<i>Barbus</i> sp.	非洲	DQ366980	DQ845860	AF180842*
	<i>Hampala macrolepidota</i>	勐腊, 云南	DQ366965	DQ845863	DQ464974*
	<i>Onychostoma gerlachi</i>	景洪, 云南	DQ366963	DQ845862	GQ406314
	<i>Onychostoma leptura</i>	西林, 广西	DQ366955	GQ406257	—
	<i>Onychostoma macrolepis</i>	泰安, 山东	DQ366942	GQ406258	—
	<i>Onychostoma sima</i>	合江, 四川	DQ366991	DQ845861	—
	<i>Percocypris pingi pingi</i>	合江, 四川	DQ366962	GQ406259	—
	<i>Puntius semifasciolatus</i>	景洪, 云南	DQ366951	GQ406260	AY856116*
	<i>Puntius conchoniuis</i>	水族馆	GQ406253	DQ845880	AY004751*
	<i>Puntius tetrazona varieties</i>	水族馆	DQ366938	EU287909	EU287909*
	<i>Sikukia stejneri</i>	勐腊, 云南	DQ366931	DQ845872	GQ406315
	<i>Sinocyclocheilus tingi</i>	抚仙湖, 云南	DQ366978	DQ845866	AY854701*
	<i>Spinibarbus hollandi</i>	屯溪, 安徽	DQ366973	DQ845865	AY195629*
	<i>Tor douronensis</i>	勐仑, 云南	DQ366945	DQ845877	DQ464986*
	<i>Tor qiaojianensis</i>	盈江, 云南	DQ366970	DQ845873	GQ406316
	<i>Tor sinensis</i>	勐腊, 云南	DQ366936	DQ845876	FJ211164*
鲤亚科	<i>Carassius auratus</i>	武汉, 湖北	DQ366941	AB006953	AB006953*
	<i>Cyprinus carpio</i>	天峨, 广西	DQ366994	X61010*	X61010*
	<i>Cyprinus multitaeniata</i>	桂平, 广西	DQ366939	DQ845845	—
	<i>Procypris rabaudi</i>	合江, 四川	DQ366969	DQ845846	GQ406317
野鲮亚科	<i>Cirrhinus molitorella</i>	藤县, 广西	DQ366959	DQ845883	AY463098*
	<i>Crossocheilus latius</i>	腾冲, 云南	DQ366982	DQ845882	—
	<i>Crossocheilus reticulatus</i>	勐仑, 云南	DQ366937	GQ406261	—
	<i>Discogobio bismargaritus</i>	柳州, 广西	DQ366947	DQ845890	GQ406318
	<i>Discogobio brachyphysallidos</i>	金秀, 广西	DQ366958	DQ845901	GQ406319
	<i>Discogobio laticeps</i>	天峨, 广西	DQ366949	DQ845889	GQ406320
	<i>Epalzeorhynchus frenatus rar</i>	水族馆	DQ366943	DQ845905	GQ406321
	<i>Garra kempfi</i>	察隅, 西藏	DQ366968	DQ845885	—
	<i>Garra orientalis</i>	乐东, 海南	DQ366957	DQ845884	GQ406322
	<i>Garra taeniata</i>	景洪, 云南	DQ366953	GQ406262	—
	<i>Henicorhynchus lineatus</i>	勐仑, 云南	DQ366935	GQ406263	GQ406323
	<i>Labeo yunnanensis</i>	勐腊, 云南	DQ366948	DQ845881	GQ406324
	<i>Lobocheilus melanotaenia</i>	勐仑, 云南	DQ366940	DQ845902	DQ464990*
	<i>Osteochilus salsburyi</i>	融安, 广西	DQ366971	DQ845892	GQ406325
	<i>Parasinilabeo assimilis</i>	融安, 广西	DQ366992	DQ845887	GQ406326
	<i>Pseudocrossocheilus bamaensis</i>	天峨, 广西	DQ366993	DQ845895	GQ406327
	<i>Ptychidio jordani</i>	天峨, 广西	DQ366974	DQ845893	GQ406328
	<i>Rectoris posehensis</i>	都安, 广西	DQ366975	DQ845891	GQ406329
	<i>Semilabeo notabilis</i>	金秀, 广西	DQ366983	DQ845886	GQ406330
	<i>Sinilabeo rendahli</i>	宜都, 湖北	DQ366932	GQ406264	—
裂腹鱼亚科	<i>Gymnocypris eckloni eckloni</i>	黄河, 青海	DQ366950	DQ845853	AY463522*
	<i>G. przewalskii przewalskii</i>	青海湖, 青海	DQ366954	DQ845851	AY463523*

表 1(续)

亚科	分类单元	采集地	GenBank 登录号		
			RAG2	16S rRNA	Cytb
雅罗鱼亚科	<i>Gymnodiptychus dybowskii</i>	伊犁, 新疆	DQ366956	DQ845859	AY463513*
	<i>Schizopygopsis younghusbandi</i> <i>younghusbandi</i>	波密, 西藏	DQ366976	GQ406265	AY463501*
	<i>Schizothorax dulongensis</i>	古永, 云南	DQ366985	DQ845849	AY954284*
	<i>Schizothorax meridionalis</i>	盈江, 云南	DQ366989	DQ845847	AY954287*
	<i>Schizothorax molesworthi</i>	察隅, 西藏	DQ366946	DQ845848	DQ126130*
	<i>Schizothorax myzostomus</i>	古永, 云南	DQ366960	DQ845850	GQ406331
	<i>Schizothorax waltoni</i>	察隅, 西藏	DQ366981	GQ406266	AY463518*
	<i>Cyprinella lutrensis</i>	GN531	DQ367019	GQ406267	AB070206*
	<i>Leuciscus leuciscus</i>	法国	DQ367007	GQ406268	AY509823*
	<i>Phoxinus phoxinus</i>	欧洲	DQ367022	GQ406269	Y10448*
	<i>Phoxinus lagowskii</i>	恒仁, 辽宁	DQ367035	GQ406270	AB162650*
	<i>Rutilus rutilus</i>	法国	DQ367003	GQ406271	AF095610*
	<i>Pimephales promelas</i>	GN530	DQ367000	GQ406272	AF117203*
	<i>Rhinichthys atratulus</i>	GN529	DQ367018	GQ406273	AF452078*
	<i>Elopichthys bambusa</i>	桃源, 湖南	DQ367016	GQ406274	GQ406332
	<i>Ochetobius elongates</i>	桃源, 湖南	DQ367012	GQ406275	AF309506*
	<i>Luciobrama macrocephalus</i>	藤县, 广西	DQ367013	GQ406276	—
	<i>Ctenopharyngodon idella</i>	横县, 广西	DQ366996	GQ406277	AF051860*
	<i>Mylopharyngodon piceus</i>	桃源, 湖南	DQ367011	GQ406278	AF051870*
	<i>Squaliobarbus curriculus</i>	武汉, 湖北	DQ367021	GQ406279	AF051877*
	<i>Tinca tinca</i>	欧洲	DQ367029	GQ406280	Y10451*
鲢亚科	<i>Aristichthys nobilis</i>	武汉, 湖北	DQ367038	GQ406281	AF051855*
	<i>Hypophthalmichthys molitrix</i>	辰溪, 湖南	DQ367002	GQ406282	AF051866*
鲴亚科	<i>Distoechodon tumirostris</i>	武汉, 湖北	DQ366998	GQ406283	AF336308*
	<i>Pseudobrama simony</i>	桃源, 湖南	DQ367028	GQ406284	AF036194*
	<i>Xenocypris argentea</i>	桃源, 湖南	DQ367024	GQ406285	AP009059*
鲮亚科	<i>Danio apogon</i>	勐腊, 云南	DQ367039	GQ406286	—
	<i>Danio rerio</i>		U71094*	AC024175*	AC024175*
	<i>Hemigrammocypripis rasborella</i>	日本	DQ367008	GQ406287	AF375863*
	<i>Nicholsicypris normalis</i>	吊罗山, 海南	DQ367034	GQ406288	—
	<i>Opsariichthys bidens</i>	桃源, 湖南	DQ367014	GQ406289	DQ367044*
	<i>Raiamas guttatus</i>	勐腊, 云南	DQ366966	GQ406290	AF051875*
	<i>Tanichthys albonubes</i>	水族馆	DQ367023	GQ406291	EF151121*
	<i>Zacco platypus</i>	金秀, 广西	DQ367010	GQ406292	AY245048*
鲃亚科	<i>Culter alburnus</i>	桃源, 湖南	DQ367004	GQ406293	AP009060*
	<i>Cultrichthys erythropterus</i>	灵山, 广西	DQ367037	GQ406299	AF051859*
	<i>Megalobrama amblycephala</i>	武汉, 湖北	DQ367025	GQ406294	AF051867*
	<i>Pseudohemiculter dispar</i>	融安, 广西	DQ367001	GQ406296	—
	<i>Pseudolaubuca sinensis</i>	桃源, 湖南	DQ367017	GQ406297	—
	<i>Rasbora lineatus</i>	横县, 广西	DQ367036	GQ406298	—
	<i>Sinibrama macrops</i>	融安, 广西	DQ367006	GQ406295	—
	<i>Toxabramis swinhonis</i>	博白, 广西	DQ367027	GQ406300	DQ464972*
鳅亚科	<i>Gobiobotia abbreviate</i>	天峨, 广西	DQ367033	GQ406301	—
	<i>Gobiobotia filifer</i>	武汉, 湖北	DQ367032	GQ406302	AY953002*
鮡亚科	<i>Abbottina rivularis</i>	南充, 四川	DQ366995	GQ406303	AF051856*
	<i>Coreius heterodon</i>	武汉, 湖北	DQ367005	GQ406304	AY953000*
	<i>Gobio gobio</i>	法国	DQ367015	GQ406305	AY426592*
	<i>Pseudogobio vaillanti</i>	天峨, 广西	DQ366999	GQ406306	AY953019*
	<i>Pseudorasbora parva</i>	金秀, 广西	DQ366997	GQ406307	AF051873*

表 1(续)

亚科	分类单元	采集地	GenBank 登录号		
			RAG2	16S rRNA	Cytb
鲮亚科	<i>Sarcocheilichthys sinensis sinensis</i>	合江, 四川	DQ367026	GQ406308	AY952983*
	<i>Saurogobio dabryi</i>	长阳, 湖北	DQ367020	GQ406309	AY245091*
	<i>Paracheilognathus meridianus</i>	横县, 广西	DQ367009	GQ406310	—
外类群	<i>Rhodeus</i> sp	西林, 广西	DQ367031	GQ406311	DQ026430*
	<i>Micronemacheilus pulcher</i>	融安, 广西	DQ367041	DQ845921	DQ105199*
	<i>Myxocyprinus asiaticus</i>	武汉, 湖北	DQ367043	DQ845896	AY986503*
	<i>Paramisgurnus dabryanus</i>	融安, 广西	DQ367040	DQ845922	AY625701*
	<i>Misgurnus</i> sp		AY804103*	AB242171*	AB242171*
	<i>Pseudogastromyzon fangi</i>	横县, 广西	DQ367042	DQ845920	DQ105221*
	<i>Gyrinocheilus</i> sp		AY804074*	AB242164*	AB242164*

a) 鲤科的亚科分类参考文献[4], *表示从 GenBank 中下载的序列, —表示缺失的数据

PCR 扩增条件如下: 94℃ 预变性 3 min; 然后是 35 个循环扩增, 每个循环包括 94℃ 变性 30 s, 45~56℃ (不同基因的退火温度有差异) 退火 30 s, 72℃ 延伸 90 s; 最后 72℃ 延伸 8 min. 扩增的 DNA 利用 0.8% 的低熔点琼脂糖凝胶电泳分离, 用 BioStar Glassmilk DNA 纯化试剂盒进行回收和纯化. 纯化后的 PCR 产物直接用于测序. 本研究分析的绝大部分序列由本课题组扩增并测序获得, 部分 16S rRNA 基因和 Cytb 基因序列从 GenBank 中获得.

对于编码蛋白质的 RAG2 基因和 Cytb 基因, 本研究采用 CLUSTALX 软件^[30]直接排序. 对于线粒体 16S rRNA 基因, 本研究先用 CLUSTALX 软件进行排序. 然后根据已有的鲤科鱼类 16S rRNA 基因的二级结构模型^[31-33]进行手工校正和排列.

1.3 数据分区和 DNA 进化模型选择

本研究采用了分区分析方法, 即针对数据不同子集采用单独的核苷酸取代模型, 并估算相应的参数, 共选择和评估了 10 种分区策略, 即从不分区到最多 8 个分区(表 2). 所有分区策略都是采用大写 P 字母加上一个表示相应分区数目的数字来表示. 不分区策略(P1)对全部序列数据都采用同一种 DNA 进化模型. 8 分区策略(P8)中, 16S rRNA 基因分为茎区和环区, RAG2 和 Cytb 基因按照蛋白质编码位点进行分区.

DNA 进化模型选择联合使用 PAUP 软件^[34]和 ModelTest 3.7 软件^[35]. 利用赤池信息量准则(akaike information criterion, AIC)^[36]来确定各分区的最佳核苷酸进化模型. ModelTest 的初始树是从基于全部序列数据得到的一组同等简约树中随机抽取^[23]. 由于

MrBayes 3.1.2 软件只允许使用包含 1, 2 和 6 种取代速率的进化模型, 因此通常不可能完全执行由 AIC 准则选择出来的模型, 只能从参数过少或过多的模型中选择. 这种情况下, 可行的办法就是选用最接近的过度参数体系模型, 从而避免参数不足可能引起的负面结果, 如枝长低估和长枝吸引^[23]. 这种方法已经为一系列模拟分析加以证实^[37,38], 分析发现, 采用过度参数体系的取代模型造成的负面效果最小, 至少普通时间可逆(general time-reversible, GTR)模型家族是这样. 本研究中, 分区贝叶斯和 ML 分析都采用 GTR+I+Γ 模型.

1.4 鲤科鱼类的系统发育分析

贝叶斯分析使用 MrBayes 3.1.2^[21]. 针对每个分区策略, 同时运行 2 个独立的马尔科夫链蒙特卡罗过程, 该过程同时运行 1 个冷链和 3 个递增的热链. 在初步分析中, 采用默认加热参数值($T=0.2$)时得到结果显示, 链间状态交换完全缺失或者不充分. 当使用 $T=0.02$ 的加热参数时, 成功进行链间状态交换的比例可以提升到 40%~80%.

马尔科夫链蒙特卡罗分析分别运行 20000000 代, 每 1000 代取样一次. 本研究采用 2 种策略来确认分析是否达到平衡状态, 首先, 利用 Tracer 程序(1.4 版)将似然值的对数、树长和全部模型参数值相对运行代数作图. 这些图形化检查分析有助于评价老化样本代数; 其次, 马尔科夫链蒙特卡罗的收敛也可以利用在线应用程序 AWTY^[37]的累积函数来作图形化的评估. 利用累积函数, 可以分析每个分支的后验概率(posterior probability, PP)值, 并且确认这些数值在所有分析中的去除老化样本之后的所有代中都是稳定

的. 一旦马尔科夫链到达平衡, 就可以预计后验概率值的稳定分布. 如果后验概率值长时间大幅度地偏离某一平衡值, 这种趋势就可以诊断为缺乏链的收敛. 本研究所做的链收敛性诊断表明, 链的收敛在每一次分析中都是在前 2000000 代之内. 本研究采取了一个更为保守的方法, 即将前 10000000 代(10000 棵抽样树)作为老化样本舍弃, 而只保留后 10000000 代(10000 棵抽样树)用于后续分析. 最后生成的 50%多数一致性树包括了平均枝长估计, 每一个节点的后验概率值, 树的可信度区间和参数估计等.

相对于不采用分区间速率变异参数(among-partition rate variation, APRV)的分区分析得到的贝叶斯树, 采用这一参数得到的树有更高的调和平均对数似然值(harmonic mean log likelihoods, HMLi). 因此在贝叶斯分区分析中, 本研究都采用“prset ratepr=variable”选项, 并且为树和序列的进化模型参数指定了一致的先验值, 为枝长指定了指数先验值.

本研究使用 RAxML 程序^[22]进行不分区和分区 ML 分析. 参照 McGuire 等人^[23]的方法, 对每一个分区策略我们都进行了 2 组分析. 第一组分析中, 估算了每一分区策略下贝叶斯树的最大似然值(用于 ML 分区策略的评价), 所用的树为 P8 分析得到的贝叶斯树. 在第二组分析中, 本研究从 200 组搜索中搜寻似然值最大的拓扑结构, 并利用 500 次取代的自展分析来衡量所揭示分支的节点支持率. 以上 2 组分析中, 本研究采纳的序列取代模型为 GTRGAMMAI.

1.5 不同分区策略的评价

本研究采用 4 种不同的判断标准来评价不同的分区策略^[23], 包括标准贝叶斯因子^[40,41], AICc^[23], 贝叶斯信息准则(bayes information criterion, BIC)^[42], 以及决定理论方法(decision theory, DT)^[43-45]. 任意一对分区模型的贝叶斯因子就是它们的边缘似然值比值. 边缘似然值通常难以计算, 因此可以近似地用 HMLi 值代替^[46]. 通过取贝叶斯因子的自然对数, 本研究计算出标准贝叶斯因子来比较任意一对分区模型. 如果贝叶斯因子大于 10($2\ln$ 贝叶斯因子 >10), 可以将其视为支持更多分区模型的证据. 由于不同分区策略得到的 HMLi 值间的关系近似于最大似然值间的关系^[23,47,48], 因此在贝叶斯框架下利用 AICc, BIC 和 DT 方法来评价分区策略时, 本研究用 HMLi 值取代最大似然值.

运用 Mrbayes 程序在固定拓扑结构的情况下进行枝长估计, 需要关闭程序的分支交换功能, 并启用节点滑块功能(需要重新设置 props). 优先考虑的分区间策略通常具有最小的 AICc, BIC 和 DT 观察值.

本研究将贝叶斯框架下利用 AICc、贝叶斯因子、BIC 和 DT 方法选择的分区策略分别同 ML 框架下利用的等级似然率检验(hierarchical likelihood-ratio tests, hLRTs), AICc, BIC 和 DT 选择的策略进行比较^[23]. 为了增加 ML 和贝叶斯策略选择流程的可比性, RAxML 和 Mrbayes 只采用了一种 DNA 取代模型(GTR+I+ Γ). 本研究分别计算不同分区策略下的 P8 贝叶斯树的最大似然值, 并将其用于 ML 框架下的分区选择标准比较. 而贝叶斯框架内的选择标准则是利用各个分区策略得到最优分支树分别进行计算(DT 除外). 为了能够在 ML 和贝叶斯分析中运用上述的分区选择标准, 我们依据现有的方法^[23]分别计算了各个分区策略参数的数量.

1.6 鲤科鱼类系统发育假设的检验

本研究中, 为了验证分区贝叶斯分析所得到的鲤科高阶元间关系的不同假设是否会被联合数据拒绝, 我们采用贝叶斯假设检验方法^[49]. 因为贝叶斯分析推断树的分布同它们的后验概率成比例, 因此, 那些通常使用的比较不同拓扑结构的统计学方法, 如近似无偏检验^[50]在贝叶斯分析中并不可行. 本研究利用 Mrbayes 软件中的 sumt 命令创建了树的 95%置信区间(在达到平衡状态以后取样), 并将这些树导入 PAUP, 然后用感兴趣的系统发育假设对树进行筛选. 如果树的 95%置信区间中有一棵或者多棵树和假设一致, 那么这个假设就不能被拒绝.

1.7 鲤科鱼类分子时间估算

采用似然速率测试方法, 评估联合数据中不同谱系间的进化速率异质性问题. 针对本研究中的联合基因序列, 我们估算得到了 GTR+I+ Γ 模型, 并将其用于有分子钟约束和无分子钟约束的系统发育分析. 结果表明, 本研究中分子钟假设被拒绝, 因为存在分子钟约束的分析和无分子钟约束的分析之间存在显著差异($P<0.005$, 自由度=107). 因此, 我们运用了 Sanderson 非参数速率平滑法(nonparametric rate smoothing, NPRS)方法^[51]来估计物种分化时间.

本研究利用 r8s 程序^[52]中的 NPRS 方法获取鲤科

鱼类的超度量树. 分歧时间估算是基于不分区(P1)策略贝叶斯分析得到的拓扑结构. 在利用 NPRS 方法估算分化时间时, 本研究使用 Powell 算法和附加惩罚函数(additive penalty function). 节点估算时间的 95%置信区间则通过长度自展分析来确定. 利用 PHYLIP 3.5c 软件包中的 SEQBOOT 命令^[53], 生成 100 个用于自展分析的伪复制数据集. 在保持树的拓扑结构不变的前提下, 采用 ML 分析估计每一个伪复制数据集中的节点长度(用于时间估计). 根据自展分析得到的 100 个年龄分布, 计算每一个节点的平均年龄.

为了精确估算物种分歧时间, 本研究采用了多个化石校正点. (1)鲤科鱼类基部节点被约束为不超过 55.8 Ma 前, 因为目前可信的最古老鲤科鱼类化石来源于始新世^[54]. (2)丁鲷属和现代雅罗鱼的分离不早于 18 Ma 前. 因为最早的丁鲷化石记录来源于中新世中期^[55,56], 此外, 欧洲淡水鱼类区系在大约 17 Ma 前~18 Ma 前之间发生了一个重要转变, 现代雅罗鱼类群开始出现, 如 *Palaeoleucisus* sp.和 *Palaeocarassius* sp. (=aff. *Abramis* sp. vel aff. *Alburnus* sp.)^[57]. (3)鲢(*Hypophthalmichthys molitrix*), 鳙(*Aristichthys nobilis*), 草鱼(*Ctenopharyngodon idella*)各自节点的物种起源不晚于 1.811 Ma 前^[58]. (4)麦穗鱼属(*Pseudorasbora*)的物种分化不晚于 3 Ma 前^[58]. (5)基于鲃属(*Barbus*)的化石记录^[10], 现代欧洲鲃(*Barbus barbus*)的物种分化时间被确定为 13 Ma 前.

2 结果

本研究共得到了 3 个基因的 4257 个排列位点, 其

中 *RAG2* 基因包括了 1287 个位点, 线粒体 *16S rRNA* 基因和 *Cytb* 基因分别包括了 1830 和 1140 个位点. 在获得的 4257 个位点中, 变异位点 2209 个, 简约性位点 1797 个. 其中, *16S rRNA* 基因的变异位点和简约性信息位点的数目分别为 190 和 106, *Cytb* 基因分别为 160 和 140, *RAG2* 基因则分别为 290 和 270.

2.1 不同分区策略对系统发育分析的影响

HMLi 和 $-\ln L$ 分别用于衡量贝叶斯和 ML 分析中数据对全部结果的解析能力. 本研究中, 简单增加分区确实导致了贝叶斯和 ML 分析中 HMLi 和 $-\ln L$ 值的大幅度提升(表 2). 这表明, 简单分区策略为联合数据提供的进化模型可能要差于复杂分区策略. 例如, 对 *RAG2* 和 *Cytb* 基因按照编码位点进行分区会对 HMLi 和 $-\ln L$ 值产生显著影响. 将 *16S rRNA* 基因进行茎和环的分区之后, HMLi 和 $-\ln L$ 较分区前分别增加了 360 和 330 个对数单位(P4 对 P3 策略). 比较包含相同分区数目的分区策略(P5a 对 P5b, P6a 对 P6b)发现, 单独对 *Cytb* 基因进行编码子分区的策略要好于仅对 *RAG2* 进行分区的策略. P8 策略分别对 *16S rRNA* 基因进行茎和环的分区和对蛋白质编码基因进行编码子分区, 因此贝叶斯和 ML 分析中分别得到了最大的 HMLi 和 $-\ln L$ 值.

本研究中, 不同分区策略的贝叶斯和 ML 分析获得的树的拓扑结构基本一致, 差异在于树中一些支持率偏低的节点($PP < 0.90$ 和自展支持率 $< 70\%$)的位置发生了变化. 所观察到的拓扑结构的最大差异在于鲮(*Danio*)在鲤科中的系统位置, 其中 P7 策略的贝叶斯分析支持其位于鲤科基部, 但支持率偏低. 不同分区策略对树长估计的影响非常有限, 同时也没有造

表 2 贝叶斯分析和最大似然分析中所使用的分区策略及其 HMLi 和 $-\ln L$

分区策略	分区	HMLi(贝叶斯)	$-\ln L$ (最大似然)
P1	所有数据联合	73918.60	73819.135
P2	mtDNA; <i>RAG2</i>	73252.40	73241.452
P3	16S; <i>Cytb</i> ; <i>RAG2</i>	72874.80	72985.832
P4	16S stems; 16S loops; <i>Cytb</i> ; <i>RAG2</i>	72510.50	72648.300
P5a	16S; <i>Cytbpos1</i> ; <i>Cytbpos2</i> ; <i>Cytbpos3</i> ; <i>RAG2</i>	72177.90	72061.933
P5b	16S; <i>Cytb</i> ; <i>RAG2pos1</i> ; <i>RAG2pos2</i> ; <i>RAG2pos3</i>	72496.10	72628.226
P6a	16S stems; 16S loops; <i>Cytbpos1</i> ; <i>Cytbpos2</i> ; <i>Cytbpos3</i> ; <i>RAG2</i>	71806.61	71717.441
P6b	16S stems; 16S loops; <i>Cytb</i> ; <i>RAG2pos1</i> ; <i>RAG2pos2</i> ; <i>RAG2pos3</i>	72130.61	72291.174
P7	16S; <i>Cytbpos1</i> ; <i>Cytbpos2</i> ; <i>Cytbpos3</i> ; <i>RAG2pos1</i> ; <i>RAG2pos2</i> ; <i>RAG2pos3</i>	71790.49	71704.833
P8	16S stems; 16S loops; <i>Cytbpos1</i> ; <i>Cytbpos2</i> ; <i>Cytbpos3</i> ; <i>RAG2pos1</i> ; <i>RAG2pos2</i> ; <i>RAG2pos3</i>	71419.06	71360.693

成 *PP*(贝叶斯分析)和自展支持率(*ML* 分析)的显著差异. 贝叶斯分析中, 内类群中得到强烈支持的节点($PP \geq 0.95$)的数量随着分区数量的增加反而有所下降. 不分区(*P1*)贝叶斯树中, 95 个内类群节点中的 73 个节点的 $PP \geq 0.95$, 而在最多分区(*P8*)贝叶斯树中, 93 个内类群节点中有 63 个节点的 $PP \geq 0.95$. 因此, 基于最复杂分区模型的贝叶斯分析在揭示鲤科中获得稳固支持的节点数目上的表现反而相对较差.

2.2 最适分区模型的选择

DT 和贝叶斯因子(对应的 *hLRT*)、*AICc* 及 *BIC* 标准分别选择出了 2 个极端的分区策略. *DT* 倾向于不分区策略(*P1*), 而其他标准优先选择的是最多分区策略(*P8*). 本研究中, 包含更多分区数目的分析策略并没有为鲤科系统发育重建提供更好的解决和支持方案, 因为所有分区策略得到的系统树在拓扑结构和节点支持率上非常相似. 因此, 在目前的鲤科系统发育分析中, 分区策略实际上并没有想象中的那么关键. 基于不分区(*P1*)策略(贝叶斯和 *ML* 分析中 *DT* 标准选择的分区模式)的系统发育分析需要估算的参数最少, 本文认为 *P1* 是贝叶斯和 *ML* 分析中的最佳策略. 尽管增加分区数目确实可以提高似然值, 但分区数目对树的拓扑结构和节点支持率的影响非常有限. 因此, 分区数目增加所导致的似然值升高可能更多地与取代模型和碱基频率参数(本文中为多余参数)的估算相关, 而与拓扑结构和枝长估算关系不大^[23].

2.3 鲤科系统发育关系

基于联合数据的贝叶斯分析为鲤科鱼类系统发育推断提供了大量信息. 所有贝叶斯分析都重现了鲤科的单系性($PP=1.0$). 基于不分区策略的贝叶斯分析就很好解决和支持了鲤科系统发育关系. 95 个内类群节点中, 75 个节点的 $PP \geq 0.95$, 2 个节点的 PP 值在 0.90 和 0.95 之间(图 1).

不分区贝叶斯分析强烈支持鲤科中几个重要分支. 然而, 鲟在鲤科中的位置问题并没有很好解决, 因为相应的节点 PP 较低(0.61), 而且 *P7* 贝叶斯分析中, 鲟位于整个鲤科的基部. 鲤系(参考文献[2])(分支 *I*)中, 野鲮鱼类的单系性(分支 *B*)以及野鲮分支与非野鲮的鲤类分支(分支 *A*, 但不包括岩原鲤(*Procypris*))间的姐妹群关系都得到了强烈支持($PP=1.0$). 但是非野鲮的鲤类分支(分支 *A*)内部单元

之间的相互关系却没有很好解决.

在雅罗鱼系(分支 *II*)中, 鲃类的单系性并没有得到支持, 但是东亚特有的鲃类, 包括鲢(*Hypophthalmichthys*)、鳙(*Aristichthys*)、草鱼(*Ctenopharyngodon*)、青鱼(*Mylopharyngodon*)、鳊(*Ochetobius*)、赤眼鳟(*Squaliobarbus*)、鳊(*Elopichthys*)、鳊(*Luciobrama*)、鲃(*Culter*)、原鲃(*Cultrichthys*)、华鲃(*Sinibrama*)、鲃(*Megalobrama*)、拟鲃(*Pseudohemiculter*)、似鲃(*Toxobramis*)、飘鱼(*Pseudolaubuca*)、圆吻鲃(*Distoechodon*)、鲃(*Xenocypris*)和似鲃(*Pseudobrama*)等, 构成一个单系类群(分支 *C*, $PP=1.0$), 拟细鲫(*Nicholsicypris*)、细鲫(*Rasborinus*)、*Hemigrammocypripis*、宽鳍鱮(*Zacco*)和馬口鱼(*Opsariichthys*)等位于分支基部. 北美和欧亚分布的雅罗鱼类共同构成一个单系类群(分支 *E*, $PP=1.0$), 但是它与丁鲃的姐妹群关系并没有获得稳固支持($PP=0.80$). 分支 *D*($PP=1.0$)包括鮡类和鲃鲃鱼类, 它与雅罗鱼类(分支 *E*)+丁鲃的姐妹群关系的 PP 为 0.60. 唐鱼(*Tanichthys*)、鲃鲃(*Rhodeus*)和副鲃(*Paraacheilognathus*)共同构成单系(分支 *F*, $PP=1.0$), 其中唐鱼为鲃鲃+副鲃分支的姐妹群. 雅罗鱼类(分支 *E*)、丁鲃、鮡类(分支 *D*)和鲃鲃类(分支 *F*)共同构成单系类群($PP=1.0$), 成为东亚鲃类(分支 *C*)的姐妹群($PP=1.0$). 基于其他 9 种分区策略得到的贝叶斯树的拓扑结构同 *P1* 贝叶斯树基本一致.

不分区策略及其他分区策略的 *ML* 分析得到的分支树和图 1 也高度相似. 主要差别在于: (1) 鲟的系统发育位置; (2) 分支 *A* 内部深层的分支关系; (3) 由雅罗鱼类、丁鲃、鮡类和鲃鲃类共同构成的分支的节点支持率. 图 1 中揭示的全部鲤科重要单系类群在 *ML* 分析中均得以重现(自展支持率 $\geq 70\%$). 唯一的例外就是非野鲮的鲤类(分支 *A*)在贝叶斯分析中的 PP 值为 0.95, 而 *ML* 分析中的自展支持率为 66%.

2.4 鲤科的物种分化时间

表 3 详细列出了利用 *r8s* 程序得到的鲤科主要节点的物种分化时间及 95% 置信区间, 这些节点在图 2 中做了标记. 利用联合数据, 本文推断鲤科在东亚出现的时间为 42.38 Ma 前(43.13 Ma 前~41.64 Ma 前). 鲤科中鲤系和雅罗鱼系的分类发生在 27.36 Ma 年前(27.84 Ma 前~26.89 Ma 前). 分化时间推算表明, 鲤系(节点 *CYN*)和雅罗鱼系(节点 *LEU*)的物种分化差不

图 1 基于不分区贝叶斯分析得到的鲤科鱼类系统发育树

每个节点的数字代表相应的后验概率。图右侧的罗马数字分别代表鲤科中的两个系群: (I), 鲤系; (II), 雅罗鱼系。标记有黑色圆点和大写黑体字母的节点分别代表了鲤科中识别出的单系类群。A: 鲤亚科; B: 野鲮亚科; C: 鲮亚科; D: 鮡亚科; E: 雅罗鱼亚科; F: 鲮亚科(包括唐鱼); G: 丁鲷亚科; H: 鲃亚科

表 3 图 2 的鲤科系统发育树中主要节点的分歧时间估算及 95% 置信区间

分支	节点	时间估算(百万年前)	
		$\bar{x} \pm SD$	95% 置信区间
鲤科, 不包括外类群	CYD	42.38 $\pm$ 3.74	41.64~43.13
鲤系	CYN	20.45 $\pm$ 1.41	20.17~20.73
鲤亚科	CYT	18.46 $\pm$ 1.04	18.25~18.67
野鲮	LAB	14.98 $\pm$ 0.95	14.79~15.17
雅罗鱼系	LEU	20.51 $\pm$ 1.22	20.27~20.76
鲮亚科+鲮亚科+雅罗鱼亚科+丁鲷	NOR	19.41 $\pm$ 0.88	19.23~19.58
雅罗鱼亚科	LES	12.02 $\pm$ 0.63	11.90~12.15
鲮亚科	GOB	14.13 $\pm$ 0.88	13.95~14.30
鲮亚科	ACH	17.11 $\pm$ 0.97	16.92~17.30
雅罗鱼亚科+丁鲷	TIN	17.90 $\pm$ 0.40	17.82~17.98
鲮亚科(东亚特有)	XEN	18.80 $\pm$ 1.12	18.58~19.02

多同时进行, 分别为 20.45 Ma 前和 20.51 Ma 前. 鲤系内部, 野鲮类(节点 LAB)的物种分化发生在 14.98 Ma 前, 而其余的鲤类(节点 CYT)物种分化时间较早(18.46 Ma 前). 而鲮鱼类(节点 ACH)、鲮类(节点 GOB)、雅罗鱼类(节点 LES)和鲮类(节点 XEN)的物种分化发生在 18.80 Ma 前到 12.02 Ma 前之间.

3 讨论

3.1 不同分区策略在系统发育分析中的表现

对于由多个基因和(或)基因区段组成的联合数据而言, 分区系统发育分析可以极大减少模型错配和系统误差. 通过对比 P1 和 P8 贝叶斯分析中每个参数的 95% 置信区间(到达平衡后取样)发现, 不同分区策略的模型参数估计存在显著异质性. 这表明, 本研究中增加分区数目可以提高贝叶斯和 ML 分析的似然值. 我们可以发现不同分区策略下的参数的置信区间互不重叠的大量例证. 基于参数估计的结果, *Cytb* 基因的编码子分区对提升 HMLi 和 $-\ln L$ 值的效果明显优于 *16S rRNA* 和 *RAG2* 基因的分区.

尽管分区对提升似然值有着重要效果, 它对拓扑结构和节点支持率的影响非常小. 本研究在贝叶斯分析中发现, 相对于不分区策略, 增加分区却导致了一些节点的 PP 值下降. 在不分区贝叶斯树中, 58 个内类群节点的 PP 值为 1.0, 而 P8 贝叶斯树中, PP 值为 1.0 的内类群节点下降为 51 个. 系统发育研究者关注的是如何在分析中采用合适的分区, 因为过少

或过多的参数设置都不利于拓扑结构和支持率的推断. 尽管改进参数设置可以减少特定分区策略下的系统误差, 但是随机误差有可能成为影响系统发育和置信度推断的重要因素. 目前, 仍不太清楚为了得到可信系统发育推论所需要的理想分区数量. 本研究中, 可以判断增加分区数目对提升 HMLi 和 $-\ln L$ 值的帮助可能更多地与改善一些多余参数(碱基频率和取代速率)的估算相关.

本研究在贝叶斯和 ML 分析中比较了 10 种分区策略, 并采用 4 种选择标准来筛选最佳分区策略. 标准贝叶斯因子 $hLRT$ 和 AICc 标准由于对附加的参数体系施加的惩罚相对较弱, 倾向选择最复杂的分区策略, 而更为严格的 BIC 和 DT 标准则分别选择了最多的和最少的分区策略. DT 方法会将相对分枝长度误差作为一个评价手段. 因此在 DT 框架下, 如果一个较少分区模型反馈的枝长估算和较多分区模型几乎相同, 那么这两个模型在系统发育推断中将没有差异. 本研究中, DT 标准选择了不分区策略, 这表明所有的分区分析并没有改善(P2~P8, 表 2)分枝长度的估算.

3.2 鲤科系统发育及其系统学

正如所料, 鲤科的单系性得以重现, 并得到了后验概率和自展支持率的强有力支持. 本研究构建了鲤科的高阶元关系框架, 揭示了其中几个获得良好支持的单系类群.

鲤科中的一个大的分支就是鲤系(分支 I, 图 1). 这个分支包含了基于传统分类学所识别的 4 个亚科的全部分类单元, 即鲮亚科、鲤亚科、野鲮亚科和裂

图 2 分歧时间估算得到的鲤科鱼类的系统发育树

超度量树的拓扑结构基于不分区贝叶斯树, 分子钟估算利用了 R8s 程序中的非参数速率平滑法. 节点的标记、平均分化时间及 95%置信区间详见表 3

腹鱼亚科^[4]。尽管分子和形态证据在鲤系分支内部的关系上还有争议,但是它们一致支持鲤系是单系。鲤系内部,目前识别的野鲮亚科的单系性得以确认,而剩余的鲃类、鲤类和裂腹鱼类(岩原鲤除外)则共同聚为 1 个分支(分支 A)。在另一包含了更多的分类单元的研究中(尚未发表),鲤系被分为两个单系类群,野鲮亚科和鲤亚科,后者包括鲃类、鲤类(包括岩原鲤)和裂腹鱼类。

鲤科中另一个主要分支是雅罗鱼系(分支 II, 图 1)。雅罗鱼系中,所有的系统发育分析都解决和支持了鲃亚科、鲴亚科、雅罗鱼亚科和东亚特有鲃亚科(Xenocyprininae)的单系性。然而,鲃亚科、鲴亚科和雅罗鱼亚科间的关系却没能有效解决,三者间的关系在不同分析中互不相同。本研究中它们三者和丁鲷共同构成一个单系群。丁鲷在鲤科中的系统位置一直被认为是分类学上的难题^[2]。不同于其他形态^[8,9]和分子^[10-12,20]数据的研究结果,本研究明确支持丁鲷同雅罗鱼类、鲃类和鲴类共同构成单系群,但仅微弱支持丁鲷和雅罗鱼类的姐妹群关系。

鲃亚科(Danioninae, 波鱼亚科, 参考 Howes^[2])鱼类的单系性在本研究中毫无疑问地遭到了拒绝。形态学上,鲃亚科实际上是一群分类单元的集合,它们因为没有办法被置于其他亚科而被划分到了鲃亚科^[2]。另外,最近的分子系统发育也明确否定鲃亚科的单系性,其中的分类单元散布于整个鲤科^[57]。因此,我们认为重新定义鲃亚科应该剔除东亚特有的一些分类单元,如鱮、马口鱼和拟细鲫等。

Chen 和 Mayden^[60]最近对鲤科(或鲤超科)鱼类的分类做了修订,他们建议识别出 10 个科(包括裸吻鱼科(Psilorhynchidae))。在这些的单系类群中,6 个类群(即丁鲷亚科、雅罗鱼亚科、鲃亚科、鲴亚科、鲃亚科和波鱼亚科)在本研究中得到了支持(图 1)。在本研究中没有包括裸吻鱼科和 Leptobarbidae 科,而鲃亚科和波鱼亚科分别对应的是本研究中的鲃亚科和鲃亚科。本研究的分子证据表明,Chen 和

Mayden 识别的鲤亚科可以被进一步划分为野鲮亚科和鲤亚科,同时,唐鱼科(Tanichthyidae)可以归并到鲴亚科。Chen 和 Mayden 将这些单系类群由亚科提升到科水平,但是我们建议保留他们在鲤科中的亚科地位。

以前的形态学研究支持将鲤科划分为 2 个系群,即有胡须的鲤类和通常胡须缺失的雅罗鱼类,但是这 2 个系群各自包括的类群及其相互关系却并不相同^[2,8,9]。然而在鲃亚科(波鱼亚科)的系统位置问题上,最近的分子系统学研究不同于传统的形态学研究。早前的形态和分子研究^[8,9,60]将鲃亚科置于雅罗鱼系,本研究的系统树也微弱支持这一结果(图 1)。另外,有研究将鲃亚科置于鲤系^[2];还有分子证据支持鲃亚科位于整个鲤科基部^[12,13,61]。在以前的分子分析中,鲃亚科系统位置的不一致可能是由于分类单元的取样不同造成的。本研究表明,鲃亚科是鲤科中的一个类群,不同于已经广为接受的鲤系和雅罗鱼系。鲃亚科属于鲤科基部类群的假设(P7 贝叶斯分析揭示)并不能被我们的联合数据拒绝。在全部树的 95%置信区间(共 18710 棵)中,由 6414 棵树同鲃亚科鱼类位于鲤科基部的假设相一致。

3.3 鲤科分支的系统发育历史

基于鲤科鱼类化石的分布,通常假设鲤科鱼类起源于始新世(Eocene)。同这个假设相一致,本研究的分子时间估算同样表明,鲤科鱼类的起源可能在始新世中期(大约四千二百万年前)。鲤科中,鲤系出现于晚渐新世(Oligocene)的早期(大约 27 Ma 前);而雅罗鱼系的起源则发生在晚渐新世(大约 26 Ma 前到 25 Ma 前之间)。

鲤系中,野鲮鱼类的物种分化发生在中中新世(Miocene)早期,鲤类的分化则在早中新世晚期。雅罗鱼系中,鲃类起源被推断发生在早中新世(约两千万年前),而鲴类、鲃类和雅罗鱼的物种分化是在中中新世时期发生(大约 18 Ma 前到 12 Ma 前)。

致谢 感谢王江新博士对本文初稿的修改和建议。

参考文献

- 1 Nelson J S, ed. Fishes of the World. New York: John Wiley and Sons Inc., 2006

- 2 Howes G J. Systematics and biogeography: An overview. In: Winfield I J, Nelson J S, eds. *Cyprinid Fishes: Systematics, Biology and Exploitation*. London: Chapman and Hall, 1991. 1–33
- 3 Hensel K. Review of the classification and of the opinions on the evolution of Cyprinoidei(Eventognathi) with an annotated list of genera and subgenera described since 1921. *Annot Zool Bot Bratislava*, 1970, 57: 1–45
- 4 陈宜瑜. 中国动物志, 硬骨鱼纲, 鲤形目(中卷). 北京: 科学出版社, 1998
- 5 Gosline W A. Unbranched dorsal-fin rays and subfamily classification in the fish family cyprinidae. *Occas Pap Mus Zool Univ Mich*, 1978, 684: 1–21
- 6 Chu Y T. Comparative studies on the scales and on the pharyngeals and their teeth in Chinese cyprinids, with particular reference to taxonomy and evolution. *Biol Bull St John's Univ(Shanghai)*, 1935, 2: 1–225
- 7 伍献文. 中国鲤科鱼类志. 上海: 上海科学技术出版社, 1964
- 8 陈湘舜, 乐佩琦, 林人端. 鲤科的科下类群及宗系发生关系. *动物分类学报*, 1984, 9: 424–440
- 9 Cavender T M, Coburn M M. Phylogenetic relationships of north American Cyprinidae. In: Mayden R L, ed. *Systematics, Historical Ecology and North American Freshwater Fishes*. Stanford, California: Stanford University Press, 1992. 293–327
- 10 Zardoya R, Doadrio I. Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *J Mol Evol*, 1999, 49: 227–237
- 11 Briolay J, Galtier N, Brito R M, et al. Molecular phylogeny of Cyprinidae inferred from cytochrome *b* DNA sequences. *Mol Phylogenet Evol*, 1998, 9: 100–108
- 12 Gilles A, Lecointre G, Faure E, et al. Mitochondrial phylogeny of the European cyprinids: implications for their systematics, reticulate evolution, and colonization time. *Mol Phylogenet Evol*, 1998, 10: 132–143
- 13 Gilles A, Lecointre G, Miquelis A, et al. Partial combination applied to phylogeny of European cyprinids using the mitochondrial control region. *Mol Phylogenet Evol*, 2001, 19: 22–33
- 14 Zardoya R, Doadrio I. Phylogenetic relationships of iberian cyprinids: systematic and biogeographical implications. *Proc R Soc B Biol Sci*, 1998, 265, 1403: 1365–1372
- 15 Durand J D, Tsigenopoulos C S, Unlu E, et al. Phylogeny and biogeography of the family Cyprinidae in the middle east inferred from cytochrome *b* DNA-evolutionary significance of this region. *Mol Phylogenet Evol*, 2002, 22: 91–100
- 16 Hanfling B, Brandl R. Phylogenetics of European cyprinids: insights from allozymes. *J Fish Biol*, 2000, 57: 265–276
- 17 Fu C Z, Wu J H, Chen J K, et al. Freshwater fish biodiversity in the Yangtze river basin of China: patterns, threats and conservation. *Biodivers Conserv*, 2003, 12, 1649–1685
- 18 Cunha C, Mesquita N, Dowling T E, et al. Phylogenetic relationships of eurasian and American cyprinids using cytochrome *b* sequences. *J Fish Biol*, 2002, 61: 929–944
- 19 He S P, Liu H Z, Chen Y Y, et al. Molecular phylogenetic relationships of eastern Asian Cyprinidae(Pisces: Cypriniformes) inferred from cytochrome *b* sequences. *Sci China Ser C-Life Sci*, 2004, 47: 130–138
- 20 Liu H Z, Chen Y Y. Phylogeny of the east Asian cyprinids inferred from sequences of the mitochondrial DNA control region. *Can J Zool*, 2003, 81: 1938–1946
- 21 Ronquist F, Huelsenbeck J P. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 2003, 19: 1572–1574
- 22 Stamatakis A. Raxml-Vi-Hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 2006, 22: 2688–2690
- 23 McGuire J A, Witt C C, Altshuler D L, et al. Phylogenetic systematics and biogeography of hummingbirds: bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Syst Biol*, 2007, 56: 837–856
- 24 Fink S V, Fink W L. Interrelationships of the Ostariophysan fishes (Teleostei). *Zool J Linn Soc*, 1981, 72: 297–353
- 25 Wang X Z, Wang J, He S P, et al. The complete mitochondrial genome of the Chinese hook snout carp *Opsariichthys bidens* (Actinopterygii: Cypriniformes) and an alternative pattern of mitogenomic evolution in vertebrate. *Gene*, 2007, 399: 11–19
- 26 Sambrook J, Fritsch E, Maniatis T, eds. *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press, 1989
- 27 Xiao W H, Zhang Y P, Liu H Z. Molecular systematics of Xenocyprinae (Teleostei: Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in east Asia. *Mol Phylogenet Evol*, 2001, 18: 163–173
- 28 Li J B, Wang X Z, Kong X H, et al. Variation patterns of the mitochondrial *16S rRNA* gene with secondary structure constraints and their application to phylogeny of cyprinine fishes (Teleostei: Cypriniformes). *Mol Phylogenet Evol*, 2008, 47: 472–487
- 29 Lovejoy N R, Collette B B. Phylogenetic relationships of new world needlefishes (Teleostei: Belontiidae) and the biogeography of transitions between marine and freshwater habitats. *Copeia*, 2001, 2: 324–338
- 30 Thompson J D, Gibson T J, Plewniak F, et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment

- aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25, 24: 4876–4882
- 31 De Rijk P, Wuyts J, Van de Peer Y, et al. The European large subunit ribosomal rna database. *Nucleic Acids Res*, 2000, 28: 177–178
 - 32 Gutell R R, Gray M W, Schnare M N. A compilation of large subunit (23s and 23s-Like) ribosomal rna structures. *Nucleic Acids Res*, 1993, 21: 3055–3074
 - 33 Gutell R R, Fox G E. A Compilation of large subunit rna sequences presented in a structural format. *Nucleic Acids Res*, 1988, 16 Suppl: r175–269
 - 34 Swofford D L. Paup*: phylogenetic analysis using parsimony (*and other methods), Version 4.0b10. 2003
 - 35 Posada D, Crandall K A. Modeltest: testing the model of DNA substitution. *Bioinformatics*, 1998, 14: 817–818
 - 36 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
 - 37 Lemmon A R, Moriarty E C. The importance of proper model assumption in bayesian phylogenetics. *Syst Biol*, 2004, 53: 265–277
 - 38 Huelsenbeck J, Rannala B. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol*, 2004, 53: 904–913
 - 39 Wilgenbusch J C, Warren D L, Swofford D L. Awty: a system for graphical exploration of mcmc convergence in bayesian phylogenetic inference. 2004. <http://ceb.csit.fsu.edu/awty>
 - 40 Kass R E, Raftery A E. Bayes factors. *J Am Stat Assoc*, 1995, 90: 773–795
 - 41 Nylander J A A, Ronquist F, Huelsenbeck J P, et al. Bayesian phylogenetic analysis of combined data. *Syst Biol*, 2004, 53: 47–67
 - 42 Schwarz G. Estimating the dimension of a model. *Ann Stat*, 1978, 6: 461–464
 - 43 Minin V, Abdo Z, Joyce P, et al. Performance-based selection of likelihood models for phylogeny estimation. *Syst Biol*, 2003, 52, 5: 674–683
 - 44 Abdo Z, Minin V N, Joyce P, et al. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. *Mol Biol Evol*, 2005, 22: 691–703
 - 45 Sullivan J, Abdo Z, Joyce P, et al. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Mol Biol Evol*, 2005, 22: 1386–1392
 - 46 Raftery A. Hypothesis testing and model selection. In: Gilks W R, Spiegelhalter D J, Richardson S, eds. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996. 163–187
 - 47 Castoe T A, Sasa M M, Parkinson C L. Modeling nucleotide evolution at the mesoscale: the phylogeny of the neotropical pitvipers of the Porthidium group (Viperidae: Crotalinae). *Mol Phylogenet Evol*, 2005, 37: 881–898
 - 48 Castoe T A, Parkinson C L. Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). *Mol Phylogenet Evol*, 2006, 39: 91–110
 - 49 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
 - 50 Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*, 2002, 51: 492–508
 - 51 Sanderson M J. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol*, 1997, 14: 1218–1231
 - 52 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
 - 53 Felsenstein J. Phylip (phylogeny inference package) version 3.5 C. 1993
 - 54 Cavender T M. The fossil record of the Cyprinidae. In: Winfield I J, Nelson J S, eds. *Cyprinid Fishes: Systematics, Biology and Exploitation*. London: Chapman and Hall, 1991. 34–54
 - 55 Hierholzer E, Mörs T. Cypriniden-Schlundzähne (Osteichthyes: Teleostei) aus dem Tertiär von Hambach (Niederrheinische Bucht, NW-Deutschland). *Palaeontographica Abteilung A*, 2003, 269: 1–38
 - 56 Sytchevskaya E. Freshwater Ichthyofauna of the Neogene of Mongolia. *Tr Sovm Sovets-Mongol Paleontol Eksped*, 1989, 39: 1–144
 - 57 Schulz-Mirbach T, Reichenbacher B. Reconstruction of oligocene and neogene freshwater fish faunas-an actualistic study on Cypriniform Otoliths. *Acta Palaeontol Pol*, 2006, 51: 283–304
 - 58 刘宪亭, 苏德造. 山西榆社盆地上新世鱼类. *古脊椎动物学报*, 1962, 6: 1–25
 - 59 Tang K L, Agnew M K, Hirt M V, et al. Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). *Mol Phylogenet Evol*, 2010, 57: 189–214
 - 60 Chen W J, Mayden R L. Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: further evidence from six nuclear genes. *Mol Phylogenet Evol*, 2009, 52: 544–549
 - 61 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, 1: 157–170