

位点特异性频率模型在昆虫分子系统发育分析中的应用^{*}

习欠云¹, 杨洁¹, 邓日强¹, 王瑾雯¹, 贾凤龙¹, 张永亮², 江青艳², 王珣章¹

(1. 中山大学生命科学学院, 广东 广州 510275;
2. 华南农业大学动物科学学院, 广东 广州 510642)

摘要: 根据位点特异性频率模型 (site specific rate model), 利用 PAUP 软件, 对线粒体 COI 基因序列不同密码子位点的数据模块分别进行了鼓甲科和水生肉食亚目在亚科或科水平上的系统发育学分析, 结果表明第二密码子数据模块获得的系统分析结果与形态学分析结果一致。表明位点特异性频率模型在线粒体 COI 基因作为标记来进行昆虫分子系统发育分析上获得了更好的效果。

关键词: 位点特异性频率模型; 鼓甲科; 线粒体 COI 基因; 系统发育; 水生肉食亚目

中图分类号: Q349⁺.5 **文献标识码:** A **文章编号:** 0529-6579 (2007) 05-0078-05

许多分子标记 (DNA) 现在被广泛地用于昆虫分子系统学的研究, 而 mtDNA 更成为研究进化的重要材料^[1]。昆虫线粒体编码基因序列, 其信息位点 (informative site) 多位于第三密码子位点, 第二密码子的信息位点最少^[2]。转换/颠换比率最高的则主要位于第一密码子^[2]。整个核苷酸序富含 A/T。核苷酸替换饱和位于第三密码子, 而第一、第二密码子的则相对未达到饱和^[3]。这些特征为位点特异性频率模型提供了理论基础^[3]。

水甲虫, 尤其是鼓甲科, 有特殊的地理分布, 对解决中国地理学划分的古老问题提供了有利的参考意义。同时一些水甲虫具有药用价值, 一些水甲虫也有食用价值, 因而研究水甲虫的分类有着一定的实际意义和科研价值。鼓甲科 (Gyrinidae) 是水生肉食亚目 (Hydradeephaga) 的一个科, 水生肉食亚目属于水生生活习性的一类鞘翅目。目前这个亚目大约有 5 500 个种, 200 多个属, 含 7 个科^[4]。鼓甲科 (Gyrinidae) 是这个亚目的第二大科^[4]。

本文根据 COI 基因序列, 利用位点特异性频率模型研究了鼓甲科的 13 个种, 涵盖了 3 个亚科的系统发育分析, 也探讨了水生肉食亚目七个科间的系统发育关系。

1 样本序列与方法

1.1 样本序列

1.1.1 35 个水生肉食亚目 (Hydradeephaga) 种的

COI 序列 其中: 鼓甲科 (Gyrinidae) 13 个, 龙虱科 (Dytiscidae) 8 个, 大卫两栖甲科 (Amphizoidae) 2 个, 沼梭科 (Halpidae) 3 个, 小粒龙虱科 (Noteridae) 4 个, 水甲科 (Hygrobiidae) 3 个, Aspidytidae 科 2 个^[4]。2 个 Ptilidae 种作为外群。其中 9 个鼓甲科的序列为本室测定, 其它序列来自 GenBank 涵盖了鼓甲科 3 个亚科和所有的属或亚属。样本序列物种名简写及序列号为: *P. acroF*, AY550882; *P. acroA*, AY550854; *H. Peltodytes* AY071816; *H. mucronatus* AY071804; *H. lineatocollis* AY071803; *A. insolens* AY071796; *A. lecontei* AY071797; *As. niobe1*, AY071808; *As. niobe2*, AY071809; *A. geotroci* AY138728; *P. princeps* AY138764; *P. obtusatus* AY138762; AY138735; *G. cinereus*; *L. lanceolatus* AY138760; *L. varius* AY071810; *M. coriaceus* AF428209; *C. schillknechti* AF428236; *Hy. australis* AY071805; *Hy. hemanni* AY071806; *Hy. maculatus* AY071807; *G. minutus* DQ266407; *G. orientalis* DQ266406; *A. striatus* AY071799; *O. villosus* AY071815; *O. fusiformis*, DQ266405; *O. ding* DQ266403; *O. Wui* DQ266402; *O. productus* DQ266404; *O. chalcus* DQ266401; *D. mellyi* DQ266408; *P. lindaisi* DQ266409; *M. IR-2001*, AY071812; *A. ellipticus* AY071798; *N. Suphisellus* AY071818; *N. Suphis* AY071817; *N.*

* 收稿日期: 2007-01-05

基金项目: 广东省自然科学基金资助项目 (04009785)

作者简介: 习欠云 (1973 年生), 男, 讲师, 博士, 现工作单位: 华南农业大学动物科学学院; 通讯联系人: 王珣章;

E-mail: wxz@zsu.edu.cn

clivicomis AY071814 *N. tenellus* AY071813

1.2 方法

1.2.1 样品预处理与总 DNA 的提取 干标本用 TE (pH 8.0) 浸泡过夜; 乙醇和福尔马林浸泡的样品用 STE 缓冲液浸泡 5 h 用冲洗缓冲液冲洗 2~3 次; 新鲜样品直接研磨。取昆虫的胸部肌肉, 剩余部分保留在 $\varphi=70\%$ 乙醇中作为标本凭证。组织剪碎后提取总 DNA (Omega 公司), DNA 保存于 2~8℃, 长期保存于 -20℃。

1.2.2 PCR 扩增与测序 引物: C1FJ2138 (Jerry) 5'-CAACATTTATTTTGA TTTTIGG-3', 和 TL2N-3014 (Pat) 5'-TCAATGCACTAATCTG-CATATTA-3', 扩增片段为 876 bp 的 COI 部分编码区。反应体系: 10× Buffer 2.0 μL; 10 mmol/L dNTP 0.4 μL; 10 μmol/L Primer 各 1.0 μL; Template DNA 1.0 μL (~10 ng); 5 U Taq DNA polymerase 0.3 μL; ddH₂O 14.3 μL 反应程序: 94℃ 5 min; 94℃ 30 s; 45℃ 30 s; 72℃ 1 min; 72℃, 8 min; 36 个循环。T 载体克隆测序。

1.2.3 序列分析

(1) 序列比对: 序列比对采用 Clustalx (version 1.83) 软件进行, 为方便分析, 除来自 Amphizoidae 科的两个种的序列 486 bp 和 692 bp 外, 将所有序列用 Bioedit 软件剪切成 744 bp 长度。序列核苷酸差异以及转换 颠换比率的统计用 MEGA vs 2.1^[5] 分析, 系统发育分析均使用 PAUP * 4.010b 软件^[6]。

(2) PAUP 法分析 (位点特异性频率模型法): 采用了 3 种方法: 最小距离法 (ME)、最大似然法 (ML) 以及最大简约法 (MP)。所有序列比对的 gap 均采用 Pairwise Deletion 原则。采用位点特异性频率模型 (site specific rate model)^[3], 即密码子位点模块法 (codon site partitioning) 来分析。该模型允许编码基因不同密码子位点拥有各自的频率。ME 法采用邻接法 (NJ) 进行, 距离矩阵按 Jukes and Cantor 的单参数法^[7]和 Kimura 双参数模式计算。系统树各分支的相对支持率采用 bootstrap 检验, 1 000 次重复取样。ML 法利用 Modeltest 2.2 对不同的核苷酸替代模式进行评价, 获得最优复杂 GTR 模型^[8]。同时使用 LRTs 和 AIC 原则构建系统发育树, 分支相对支持率采用 bootstrap 检验^[9], 100 次重复取样。MP 法针对各密码子的饱和情况采用不同的加权值加权或不加权处理。分支相对支持率采用 bootstrap 检验, 1 000 次重复取样。同时采用 Autodecay V5.0 辅助检验分支支持率。

2 结 果

2.1 COI 序列特点及三个密码子位点的核苷酸替换饱和性分析

本文构建了两个序列数据矩阵, 一个 15 个序列数据矩阵 (含 13 个豉甲科和 2 个外围群序列), 一个是 37 个序列数据矩阵 (含 35 个水生肉食亚目和 2 个外围群序列), 分别对豉甲科和水生肉食亚目的系统发育进行分析。利用 MEGA vs 2.1 软件^[5]分析两个数据矩阵的序列结构, 结果发现在 15 个序列数据矩阵中有 228 个信息位点 (informative site), 占总位点数的 30.64%, 其中第一密码子位点 47 个, 占总位点数的 6.32%, 第二密码子位点 16 个, 占总位点数的 2.15%, 第三密码子位点 165 个, 占总位点数的 22.17%。所有位点的转换 颠换比率为 0.79。在 37 个序列数据矩阵中, 有 282 个信息位点 (informative site), 占总位点数的 37.90%, 其中第一密码子位点 65 个, 占总位点数的 8.74%, 第二密码子位点 22 个, 占总位点数的 2.96%, 第三密码子位点 194 个, 占总位点数的 26.07%。所有位点的转换 颠换比率为 0.71。揭示了 COI 基因在第三密码子的信息位点最丰富, 分别为 165 或 194 个。而转换 颠换比率则主要位于第一密码子, 分别为 1.83 或 2.30。整个核苷酸序富含 A/T, A+T 含量分别为 70.7% 和 71.7%, 而且在第三密码子位点高达 90.0%, 这是昆虫线粒体基因的共同特征^[2]。此外, 通过未校正的 p 距离值与 Γ 分布的 Kimura 双参数距离值的对应比较, 获得各密码子位点的核苷酸替换饱和情况, 高饱和度位于第三密码子, 而第一、第二密码子的则相对未达到饱和。

2.2 豉甲科系统发育分析

采用位点特异性频率模型将 15 个 COI 序列数据矩阵划分为不同的数据模块, 包括所有密码子数据矩阵、第一密码子数据矩阵、第二密码子数据矩阵, 以及第一、第二密码子数据矩阵。

在 MP 法分析中, 分析结果表明最优树的获得来自第二密码子数据矩阵, 树长 length=465 steps, CI=0.51, RI=0.49。此外, 所有排除第三密码子位点数据的严格一致树均表现相似的拓扑结构, 但包括了第三密码子位点数据所构建的树中主要表现为 Enhydriinae 和 Gyrininae 亚科内分类的紊乱, 表明第三密码子位点数据对树的构建有较大的影响。在来自第二密码子数据矩阵的最优树中, 除 Gyrininae 亚科恢复成一个单系外, 另外两个亚科则为并系。而并系的 Orectochilinae 亚科很好地形成了

于树的基部, 并且 Gyrinidae 形成了较高支持 (bootstrap=65%) 的单系。在 Gyrinidae 中, 3 个亚科呈现层次性排列 ((Gyrinae + Enhydrinae) + Orectochilinae), 与 15 序列矩阵获得的最优树结果一致。此外形成了 Dytiscoidea 总科 (=Dytiscidae, Hygrobiidae, Noteridae and Amphizoidae), 并恢复成一单系。Hygrobiidae 和 Noteridae 形成姐妹群位于 Dytiscoidea 总科的基部。Aspidytidae 与 Amphizoidae + Dytiscidae 形成姐妹群, Amphizoidae 与 Dytiscidae 形成姐妹群。

3 讨 论

研究表明用位点特异性频率信息模块来进行系统分析往往比用整个数据矩阵信息有效得多^[12-14]。本研究利用不同密码子位点的信息数据模块来分析 Gyrinidae 科以及 Hydradephaga 亚目的系统发育关系, 结果显示 PAUP 法对密码子数据模块的分析得到了理想的结果。虽然第二密码子含有限的信息位点, 比其它两个密码子的信息位点少, 但是根据第二密码子的信息模块所构建的 ME、MP 和 ML 树却比其他密码子数据模块, 包括全部密码子的信息模块更合理。

在 Gyrinidae 科的系统发育分析中, 来自第二密码子数据模块所构建的大多数树 (MP、ME 和 ML) 获得了最优的拓扑结构 (图 1), 在 3 个亚科中, 亚科 Orectochilinae 和 Gyrinae 以高的支持率获得了单系, 但亚科 Enhydrinae 没有形成单系。此外, 来自亚科 Orectochilinae 的 *Orectochilus dinghushanensis* 是一个新的种 (资料没发表), 其成功地分类于亚属 *Patrus* 中。在所有获得的树中, 种 *Pororhynchus landa is landa is* 始终位于树的基部, 如果不考虑这个种, 那么其它 Gyrinidae 的种得到了与形态分类相同的拓扑结构^[10], 也就是根据 COI 基因序列而生的分子系统学分类与形态学分类是一致的。结果表明 COI 基因对研究 Gyrinidae 科的系统发育分类是一个有效的分子标记, 也说明第二密码子在分类上含有充分的系统发育信息。而第三密码子的饱和性以及第一密码子的高转换率也许是影响分类紊乱的原因, 或者利用更复杂的替换模型可能克服这种缺点。种 *Pororhynchus landa is landa is* 在树中的异常位置为该种分类研究也许提供了一个参考。

虽然 Hydradephaga 亚目被认为一个单系, 但其科间关系却一直未得到共识。目前普遍的观点就是 Gyrinidae 科位于 Hydradephaga 亚目内的基础位置。Kavanaugh^[15] 没有解决 Dytiscoidea 总科分化枝

问题。Beutel 根据成虫或幼虫的形态特征解决了 Dytiscoidea 总科分化枝问题, 把由 Noteridae, Amphizoidae, Hygrobiidae 和 Dytiscidae 组成的 Dytiscoidea 总科恢复成一单系。Ribera^[11] 首次根据 18S rRNA 的序列信息获得了相同的结果, 只是 Hygrobiidae 科被分类于 Dytiscidae 内, 而不是与之成为姐妹关系。Balke^[12] 结合形态学和 18S, 16S, 12S, H3 和 COI 分子标记而获得的分类树中同样获得了 Dytiscoidea 总科, 其中 Hygrobiidae 位于分化枝 (Amphizoidae + Dytiscidae) 的外面。本文研究结论与 Balke^[12] 一致 (图 2)。

另一个有争议的问题是原始的 Hydradephaga 亚目是来自水生还是陆生? 现在普遍认为原始的 Hydradephaga 亚目是来自古老陆生生物^[16]。根据其入侵水生环境的顺序, 一个观点认为 Gyrinidae 是第一个入侵的, 接下来的两次入侵分别是 Haliplidae 的祖先和 Dytiscoidea 总科 (=Dytiscidae, Hygrobiidae, Noteridae 和 Amphizoidae) 的祖先。根据这一理论, Hydradephaga 应该呈现三个分化枝。本文研究也正好与此一致 (图 2), 该结果也同时反映了三个起源的系统发育关系。本文研究结果也与当前的系统分类结果^[11, 12, 17] 相一致, Hydradephaga 是一个高支持的单系, 也排除了原先形态学认为的 Hydradephaga 是一个多起源的假说。

参考文献:

- [1] 张亚平, 施立明. 动物线粒体 DNA 多态性的研究概况 [J]. 动物学研究, 1992, 13: 289-298
- [2] LI HM, DENG RQ, WANG JW, et al. A preliminary phylogeny of the Pentatomomorpha (Hemiptera: Heteroptera) based on nuclear 18S rDNA and mitochondrial DNA sequences [J]. Mol Phylogenet Evol, 2005, 37: 313-326
- [3] KJER K M, BLAHL K R J, HOLZENTHAL R W. Phylogeny of Trichoptera (Caddisflies): characterization of signal and noise within multiple datasets [J]. Syst Biol, 2001, 50: 781-816
- [4] RIBERA J, BEUTEL R G, BALKE M, et al. Discovery of Aspidytidae: a new family of aquatic Coleoptera [J]. Proceedings of the Royal Society of London, Ser. B, 2002, 269: 2351-2356
- [5] KUMAR M, TAMURA K, KJER K M, et al. MEGA2: Molecular Evolutionary Genetics Analysis Software [CP]. Arizona State University, Tempe, AZ, USA, 2001.
- [6] SWOFFORD D L, OLSEN J. Phylogenetic Analysis Using Parsimony (* and Other Methods) [CP]. Version 4.0b10. Sinauer Associates, Inc. Publishers, Sunderland,

- Massachusetts 2002
- [7] JUKES T H, CANTOR C R. Evolution of protein molecules [C] //MUNRO H N, ed. Mammalian Protein Metabolism. New York: Academic Press, 1969: 21 – 132
- [8] RODRIGUEZ P. Minority adoptions and agency practices [J]. Social Work, 1990, 35(6): 528 – 531.
- [9] FELSENSTEIN J. Confidence limits on phylogenies: an approach using the bootstrap [J]. Evolution, 1985, 39: 783 – 791.
- [10] BEUTEL R G. Phylogenetic analysis of the family Gyrinidae (Coleoptera) based on meso – and meta thoracic characters [J]. Quaestiones Entomologicae, 1990, 26: 163 – 191.
- [11] RIBERA J, JAMES E, HOGAN, et al. Phylogeny of Hydradeephagan Water Beetles Inferred from 18S rRNA Sequences [J]. Molecular Phylogenetics and Evolution, 2002, 23(1): 43 – 62
- [12] BALKE M, RIBERA J, BEUTEL R G. The systematic position of Aspidytidae: the diversification of Dytiscoidea (Coleoptera: Hydradeephaga) and the phylogenetic signal of third codon positions [J]. Blackwell Verlag, Berlin, JZS, 2005, 43(3): 223 – 242
- [13] BRANDLEY M B, SCHMITZ A, REEDER T W. Partitioned Bayesian analyses: partition choice and the phylogenetic relationships of scincid lizards [J]. Syst Biol, 2005, 54(3): 373 – 390
- [14] NYLANDER J A A, RONQUIST F, HUELSENBECK J P, et al. Bayesian Phylogenetic Analysis of Combined Data [J]. Sys Biol, 2004, 53: 47 – 67.
- [15] KAVANAUGH D H. A systematic review of amphizoid beetles (Amphizoidae: Coleoptera) and their phylogenetic relationships to other Adephaga [J]. Proceedings of the California Academy of Sciences, 1986, 44(6): 67 – 109
- [16] LAWRENCE J F, NEWTON A F. Evolution and classification of beetles [J]. Annual Review of Ecology and Systematics, 1982, 13: 261 – 290
- [17] SHULL V L, VOGLER A P, BAKER M D, et al. Sequence alignment of 18S ribosomal RNA and the basal relationships of Hydradeephagan beetles: Evidence for monophyly of aquatic families and the placement of Tychypachidae [J]. Syst Biol, 2001, 50(6): 945 – 969.

A Case Study Using Site-specific Rate Model in Insect Phylogenetic Tree Reconstruction

XIQian-yun¹, YANG Jie¹, DENG Riqiang¹, WANG Jin-wen¹, JIA Feng-bing¹,
ZHANG Yong-liang², JIANG Qing-yan², WANG Xun-zhang¹

(1. State Key Laboratory for Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

2. College of Animal Science, South China Agricultural University, Guangzhou 510642, China)

Abstract Comparison of different codon partition data schemes in phylogenetic tree construction demonstrated the advantage of second codon partition data in phylogenetic analyses at subfamily level for Gyrinidae and at family level for Hydradeephaga using PAUP or MrBayes with ME, MP, or ML criteria based on partial mitochondrial COI gene sequences. All optimal trees inferred using PAUP were obtained based on second codon partition data. Trees inferred from morphological data supported the optimal trees. The results suggest the site-specific rate model (codon-partitioning schemes) is available approach to reconstruct insect phylogenetic trees based on partial mitochondrial COI gene sequences.

Key words site-specific rate model; Gyrinidae; mitochondrial COI gene; phylogenetic relationship; Hydradeephaga