Molecular Evolution of miR-34 Gene Family

GAO Jia-li, LUO Yu-ping*, LI Si-guang

( College of Life Sciences, Nanchang University, Nanchang, 330031, China )

Abstract: miR-34 gene can regulate early development of mammalian. Using bioinformatics analysis of sequenced metazoan genomes, 54 miR-34 genes of 33 different animal species were found. Eighteen miR-34 genes were identified in the first time, suggesting miRNA are highly conservative and widespread in Eukaryotes. Of these miR-34 gene homologues identified in metazoan, 80% locate in the intergenic region, the others are in the intron or 3'UTR of protein-coding gene. miR-34 gene is conserved in various animals, and the mature sequences and precursors of miR-34 gene in metazoan analyzed share 68% and 38.89% sequence identity, respectively. However there is only a miR-34 gene in invertebrate and three miR-34 genes, miR-34a, miR-34b and miR-34c are present in almost all vertebrate. The phylogenetic analysis reveals that the miR-34 gene family of vertebrate originated from local and tandem duplications followed by mutation of individual base after divergence from invertebrate.

Key words: microRNA; miR-34; Molecular evolution
2004)，且其大部分是位于内含子区（Kim & Nam，2006）。一些内含子 miRNA 基因的位置在
不同的物种中是高度保守的，例如，miR-7 在哺乳动物和昆虫的 hnRNP K 蛋白内含子区都有发现。另
外一个很有趣的是，miR-106b 基因家族在人类和老鼠中均位于 MCM7 基因的第 13 个内含子上
（Kim & Nam，2006）。miRNA 不仅在基因位置上保守，序列上也呈现出高度的同源性（Pasquinelli
et al., 2000; Ruvkun et al., 2001; Lee & Ambros, 2001）。在肝虫 Caenorhabditis elegans 中所发现的
miRNA，85% 都可以在 C. briggsae 的基因组中找到同源序列（Lau et al., 2001）。植物中，miR-39 的
同源序列在拟南芥、水稻和烟草中都存在（Reinhart et al., 2002; Llave et al., 2002）。

miRNA 高度的保守性与功能的重要性有着密切的关系。Houbaviy et al.（2003）研究发现，
miR-34 基因在小鼠胚胎干细胞中特异性表达，它们可能参与维持胚胎细胞的全能性以及哺乳动物的
早期发育调控。miRNA 与其靶基因的进化有着密切的联系，研究其进化历史有助于进一步了解其作
用机制和功能。本文依据 miRNA 基因在物种间的保守性，利用生物信息学同源性搜寻、比对的方法
分析 miR-34 基因家族的系统进化历史。

1 材料和方法

1.1 数据的获得

从 miRNA 数据库（http://microrna.sanger.ac.uk/cgi-bin/sequences/browse.pl）中获得人 mir-34 基因的
前体序列（has-mir-34a, has-mir-34b, has-mir-34c）以及线虫 mir-34 基因的前体序列（cel-mir-34）。

1.2 同源性搜寻

分别用人和线虫的 miR-34 基因的前体序列通过同源性比对的方法在所有动物物种中进行搜寻。本
数据库（http://www.ensembl.org/index.html）以及 UCSC 数据库（http://genome.ucsc.edu/cgi-bin/
hgBlat）。在前两个中是通过 BlastN（Altschul et al., 1990）程序进行搜寻，在 UCSC 数据库中是通过
BLAT（Kent, 2002）程序搜寻。

1.3 二级结构预测

利用在线的 Mfold 程序（http://www.bioinfo.rpi.edu/applications/mfold/ma/form1.cgi）（Zuker,
2003）对所有搜寻到的新序列进行二级结构的预测。将预测结果与 miRNA 的鉴定标准（Ambros et
al., 2003）作比较，成熟序列长度应为 22 核苷酸，计算机模拟可以得到 miRNA 前体保守的二级
结构，并且成熟序列应位于发夹结构茎区的 5’或 3’侧。通过序列的筛选，最终确定 miR-34 基因的
同源序列。

1.4 同源性比对

采用 Clustal W 程序（Thompson et al., 1994）对 mir-34 基因的所有同源序列进行多序列比对
（multiple sequence alignment，MSA）。

1.5 构建系统发育树

最后，用 MEGA（3.1 版本）软件对搜寻到的 mir-34 基因及其同源序列进行多序列比对，并做出相
应的系统进化树。

2 结果分析

2.1 miR-34 基因家族新成员的鉴定及特征分析

通过同源性搜寻的方法，分别在 33 个不同的动物物种中，找到了 54 条 Mir-34 基因的同源序列
（Tab. 1），其中 18 条为新发现的序列（斜体），它们都具有典型的茎环状二级结构（Fig. 1）。本文
中一致用 mir-34a 代表 miRNA，mir-34b（斜体）代表 miRNA 前体基因。mir-34 基因及其同系物广泛存
在于节肢动物、线虫纲动物以及许多哺乳动物中，保守性很高。mir-34 家族中大部分成员（70%）
的基因位置已知，其中 84% 都位于基因间隔区（intergenic region, IGR），少数位于蛋白编码基因的
内含子区和 3’UTR 上。狗 mir-34a 基因位于其 5 号染色体 LOC608460 基因的内含子区，原鸡 mir-34a
基因位于 ENSGALT00000035086/87 基因的内含子区。此外，人和黑猩猩的 mir-34b 和 mir-34c 基因
分别位于同源基因的内含子区和 3’UTR 上。

序列分析发现在人、黑猩猩、牛、猪、鼠、大鼠、斑马鱼以及原鸡 8 个物种中，mir-34b 基因与
mir-34c 基因均紧密相连，位于同一染色体上。人和黑猩猩的 mir-34b 和 mir-34c 基因分别位于同
源基因的内含子区和 3’UTR 上，而在其他物种中 mir-34b 基因和 mir-34c 基因都位于同一基因间隔
区内（Fig. 2）。这一组织形式的特征暗示了这两个基因（mir-34b 与 mir-34c）很可能是在进化过程
中由串联重复产生的（Tanzer & Stadler, 2004）。

2.2 miR-34 基因家族同源性分析
<table>
<thead>
<tr>
<th>miR 名称</th>
<th>物种 Species</th>
<th>基因位置</th>
<th>基因组织形式</th>
<th>正/负链</th>
</tr>
</thead>
<tbody>
<tr>
<td>ago-miR-34</td>
<td>Anopheles gambiae</td>
<td>chr2R: 28232717–28232737</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>age-miR-34</td>
<td>Aedes aegypti</td>
<td>未知</td>
<td>未知</td>
<td>未知</td>
</tr>
<tr>
<td>ame-miR-34</td>
<td>Bombus mori</td>
<td>BAA01147222: 1416–1513</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>bao-miR-34</td>
<td>Bos taurus</td>
<td>chr16: 30629167–30629149</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>bta-miR-34</td>
<td>B. taurus</td>
<td>Un.: 19678–19699</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>bta-miR-34c</td>
<td>B. taurus</td>
<td>Un.: 19878–19899</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>chr-miR-34</td>
<td>Caenorhabditis briggsae</td>
<td>chr11: 35197686–35197707</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>cel-miR-34</td>
<td>C. elegans</td>
<td>chrX: 2969803–2969824</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>cfa-miR-34</td>
<td>Canis familiaris</td>
<td>chr5: 65484449–65484471</td>
<td>LOC608460 内含子区</td>
<td>未知</td>
</tr>
<tr>
<td>cfa-miR-34b</td>
<td>C. familiaris</td>
<td>chr5: 24566992–24567014</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>cfa-miR-34c</td>
<td>D. melanogaster</td>
<td>chr3R: 5926674–5926695</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>dan-miR-34</td>
<td>Drosophila melanogaster</td>
<td>scaffold13340: 6488582–6488602</td>
<td>未知</td>
<td>未知</td>
</tr>
<tr>
<td>der-miR-34</td>
<td>Drosophila pseudobscura</td>
<td>scaffold13340: 64600166–64600208</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>dgr-miR-34</td>
<td>D. simulans</td>
<td>scaffold3257: 1475912–1476028</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>dve-miR-34</td>
<td>D. viridis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>dve-miR-34c</td>
<td>D. viridis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>dve-miR-34d</td>
<td>D. viridis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>eae-miR-34</td>
<td>Echinops telfairi</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>ena-miR-34</td>
<td>Gallus gallus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>hsa-miR-34</td>
<td>Homo sapiens</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>hsa-miR-34c</td>
<td>H. sapiens</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>pti-miR-34</td>
<td>Pan troglodytes</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>pta-miR-34c</td>
<td>P. troglodytes</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>lla-miR-34</td>
<td>Lagochroa lagochroa</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>mml-miR-34a</td>
<td>Masochoa maclatika</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>ame-miR-34</td>
<td>M. nemestrina</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>gep-miR-34</td>
<td>Pan paniscus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>pyi-miR-34</td>
<td>Pongo pygmaeus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>mmu-miR-34</td>
<td>Mus musculus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>mmu-miR-34a</td>
<td>M. musculus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>mmu-miR-34b</td>
<td>M. musculus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>rno-miR-34</td>
<td>Rattus norvegicus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>rno-miR-34a</td>
<td>R. norvegicus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>rno-miR-34b</td>
<td>R. norvegicus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>slo-miR-34</td>
<td>Saguinus labiatus</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>xtr-miR-34a</td>
<td>Xenopus tropicalis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>xtr-miR-34b-a</td>
<td>X. tropicalis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>xtr-miR-34b-2a</td>
<td>X. tropicalis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>xtr-miR-34b-3a</td>
<td>X. tropicalis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
<tr>
<td>xtr-miR-34b-3b</td>
<td>X. tropicalis</td>
<td>scaffold13340: 64600496–64600968</td>
<td>基因间区隔</td>
<td>未知</td>
</tr>
</tbody>
</table>

▲表示使用NCBI数据库搜寻获得，*表示使用UCSC数据库，●表示使用ENSEMBL数据库。斜体表示新发现的miR-34同源序列。

Table 1: miR-34 and its orthologue found in sequenced metazoan.
图 1 新 Mir-34 基因家族成员的三级结构预测
Fig. 1 Predicted hairpin secondary structures of identified novel miRNAs using RNAviz software.
成熟 microRNA 序列用下划线表示（Mature miRNA is underlined）。

图 2 八个物种中 mir-34b 和 mir-34c 基因位置示意图
Fig. 2 Characterization of genomic organization of mir-34b and mir-34c in the eight animal species.
方框表示外显子，实线箭头表示转录方向，空心箭头分别表示 mir-34b、mir-34c 基因及其上下游基因，虚线部分表示未知序列。
Exons are represented by boxes, mir-34b, mir-34c and neighbor gene sequences are represented by arrow. The cartoon is not drawn to scale.
miR-34 家族成员的成熟序列同源性很高，约为 68% (Tab. 2)。序列长度是相近的，22—23 nt，保守的碱基数为 19 个。miRNA 的 5’端的 2—8 个核苷酸对于 miRNA 与靶序列 3’ UTR 区的配对非常重要，因而被称为种子序列（Doech & Sharp, 2004）。这段序列在同源的 miRNA 中通常是保守的，比对的结果正证明了这一点。从差异碱基的变化特征来看，miR-34 基因及其同系物在进化过程中很可能发生了碱基的缺失。例如 miR-34a 很可能在 11 位缺失了一个碱基。此外，10—11 位碱基在四个家族成员中保守性相对不高，这可能是由于两个碱基配对的结构更有利于 miRNA 与其靶序列的配对（Doech & Sharp, 2004）。

多序列比对分析显示 miR-34 基因家族成员前体基因同源性约为 38.89% (Fig. 3)。图中星号部
分表示了同源性较高的碱基。有少数位点的碱基呈现出类似于单核苷酸多态性 (SNP) 的现象（Tong et al., 2006）。

<table>
<thead>
<tr>
<th>表 2</th>
<th>MIR-34 基因家族成员成熟序列比较</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIR 名称</td>
<td>成熟序列</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>miR-34</td>
<td>uggagagtuggaguugguu</td>
</tr>
<tr>
<td>miR-34a</td>
<td>uggagagtuggaguugguu</td>
</tr>
<tr>
<td>miR-34b</td>
<td>uggagagtuggaguugguu</td>
</tr>
<tr>
<td>miR-34c</td>
<td>uggagagtuggaguugguu</td>
</tr>
</tbody>
</table>

保守碱基用阴影标注，缺失的碱基用短横线 (-) 表示。 Highly conserved bases are shaded. Hyphen denotes base deletion.
2.3 miR-34系系发育分析

搜寻到的所有miR-34同源序列分别分布在节肢动物门、线虫纲、鱼纲、两栖纲、鸟纲和哺乳动物中。选择这些纲目中系统发育研究的模式物种用MEGA 3.1软件构建系统进化树（Jing et al., 2004; Sempere et al., 2006）（Fig. 4）。总的来看，进化树大致可分为三支：mir-34b和mir-34c位于第一分支中；其次是mir-34a；第三分支中是mir-34。从进化树的分支和距离上来看，mir-34基因的分支出现在进化早期并且一直延续到现在，因而它很可能是miR-34基因家族中最早出现的基因形式，即祖先基因。mir-34a基因在所有被建树研究的哺乳动物（人、黑猩猩、小鼠、大鼠、牛、狗）以及鸟类（原鸡）、两栖纲动物（非洲爪蟾）中都保守，这一分支中只有一个例外是dre-mir-34c。正如前面基因位置分析结果那样，mir-34b基因与mir-34c基因亲缘关系很近，并且它们与mir-34a基因是平行进化的。

在无脊椎动物中，miR-34基因只有一个拷贝（mir-34），而在脊椎动物中miR-34基因有3个拷贝（mir-34a, mir-34b, mir-34c），且miR-34基因家族的3个成员稳定地存在于从鱼类到哺乳动物中。基于以上的鉴定和分析结果，可推测出miR-34基因家族可能的系统进化历史（Fig. 5）。miR-34基因作为该家族最古老的基因形式仅存在于节肢动物和线虫纲动物等无脊椎动物中。在由无脊椎向脊椎
图 5 miR-34 基因家族可能的系统进化历史

Fig. 5 A plausible scenario for the evolution of the miR-34 gene family

3 讨 论

MicroRNA 作为生物体正常表达的内源非编码 RNA 分子，具有高度的保守性。Tanzer & Sadler (2004) 在对 miR-17 基因簇的分子进化研究时发现，该基因簇进化可能与脊椎动物早期进化有关，进化过程伴随着个别基因的复制和缺失以及整个基因簇的复制。此外，水稻 miR-395 基因家族的分子进化研究结果也得出了相似的结论，基因家族的形成通常被认为是基因复制的结果（Elemento et al., 2002）。目前，在已测序的 33 种不同的动物中发现了 54 个 miR-34 基因。表明 miR-34 是高度保守的，在广泛存在于后生动物中。在无脊椎动物和脊椎动物中 miR-34 基因家族成员个数并不相同，表明与 miR-17 基因簇和 miR-395 基因家族类似，miR-34 基因在从无脊椎动物到脊椎动物的进化过程中也发生了基因的串联重复和局部重复（Fig. 5）。在漫长的进化过程中，3 个成员组成的 miR-34 基因家族保守地存在于几乎所有的脊椎动物中，这是否意味着与无脊椎动物相比，脊椎动物需要更多的 mir-34 基因来形成一个基因调控网络，从而成功的行使其调控功能；或者这些 miR-34 基因分别在脊椎动物的不同组织中特异性表达。

进化的现象。非编码小 RNA 基因的进化有共通之处，miRNA 的进化也主要是通过重复及随后的序列变异来完成的（Pasquinielli & Ruvkun, 2002；Grun et al., 2005；Houbaviy et al., 2005）。由于 miRNA 行使其功能是通过与其靶序列互补配对的方式而完成，它很可能也是与其靶基因共进化的（Allen et al., 2004）。对非编码小 RNA 基因的系统进化分析将有助于进一步了解其起源和作用机制。

参考文献：


