*miRNA-9*基因家族进化分析及其靶基因预测

罗丹丹 罗玉萍 彭 娟 李思光*
(南昌大学生命科学与食品工程学院, 南昌 330031)

摘要 microRNAs(miRNA)是真核生物中一类长度约为21~25个核苷酸的非编码小分子RNA, 在转录后水平调控基因的表达。该文在miRBase中搜索后生动物的*mir-9*基因序列。47个物种中共搜索到120条*mir-9*基因序列, 说明*mir-9*基因家族广泛存在于不同物种中。基因定位显示86%的*mir-9*基因存在于基因间隔区(IGR), 多序列比对发现*miR-9*基因家族成熟序列的第2位到第8位碱基以及第14位到第18位碱基为保守碱基。进化分析表明*mir-9b*和*mir-9c*可能是此基因家族最早出现的基因形式, 即祖先基因。这些祖先基因经过串联重复、大片段重复、个别碱基的缺失及突变等方式形成了脊椎动物中*miR-9-1*至*miR-9-7*数个基因。分别采用四个miRNA靶基因预测软件对*mmu*-miR-9的靶基因进行预测, 发现miR-9与神经系统发育、心肌系统疾病和跨膜运输系统等密切相关。该研究为今后进一步研究miRNA调控的神经系统发生和神经细胞生长与分化的机制奠定了基础。

关键词 *miRNA-9*; 进化; 靶基因

miRNAs是一类长度为20~28 nt的非编码小分子RNA, 通过与靶mRNA的互补配对而在转录后水平上负调控基因的表达[1]。miRNA基因以单拷贝、多拷贝和基因簇等多种形式存在于基因组中。成簇基因之间有的具有同源性, 有的无同源性, 其中串联基因簇(tandem cluster)中的miRNA一般呈现为协同表达[2,3], 暗示着miRNA基因存在复杂的调控机制。早期的研究认为miRNA基因主要存在于基因间隔区, 但近期的研究发现大部分哺乳动物的miRNA基因主要定位在转录单位(transcript units, TUs)[2,4], 且大多数位于内含子区。

不同物种间的同源miRNA存在高度进化保守性。如miR-1、miR-34、miR-87在无脊椎动物和脊椎动物中高度保守, 序列比对分析发现这些保守序列的碱基差异仅为1-2 nt。Lau等[3]在美丽新小杆线虫(*Caenorhabditis elegans*)中发现的miRNA, 85%在*C. briggsaze*基因组序列中具有同源序列。miRNA在物种间的保守性使其成为探索非编码RNA进化规律的理想分子。

已有研究报道, 小鼠和人等哺乳动物中miR-9主要在神经组织中表达[5], 并通过调控*BAF53a*[6]、*stathmin*[7]等基因的表达参与神经细胞的增殖、迁移等活动。近期研究显示miR-9还控制着乳腺肿瘤的扩散和发展[8,9], 所以, 对miR-9的深入研究不仅利于揭示其所介导的调控网络, 还将为人类疾病的诊断和治疗提供全新的视野。本文根据miRNA的进化保守性, 通过多序列比对及构建系统进化树的方式分析*mir-9*基因家族的进化历史。采用TargetScan、PicTar、microRNA.org和MicroCosm Targrts等四个miRNA靶基因在线分析软件分析miR-9的靶基因。通过The Gene Ontology、Kyoto Encyclopedia of Genes and Genomes和David软件分析了靶基因的生物学功能, 进一步了解miR-9在生物体内的角色及其可能的作用机制。

1 材料与方法

1.1 *mir-9*序列的获得

从miRBase(http://www.mirbase.org/cgi-bin/browse.pl)中下载部分后生动物的*mir-9*基因前体序列及成熟序列。

1.2 几种模式生物的*mir-9*基因定位

采用NCBI数据库BLAST(http://blast.ncbi.nlm.

收稿日期: 2010-10-11 接受日期: 2011-02-25
国家自然科学基金(No.30971473, No.31071138)资助项目
*通讯作者。Tel: 0791- 8304099, E-mail: siguangli@163.com

nih.gov/Blast.cgi)程序, 分别用人、黑猩猩、牛、小鼠、斑马鱼、原鸡和黑腹果蝇等7个模式物种的miRNA前体序列对其基因组数据库进行搜索(BLAST Assembled RefSeq Genomes), 以确定*miR-9*在基因组中的位置。

1.3 系统发生分析

用Clustal X 1.83软件对所有已搜索物种的*mir-9*基因的前体序列进行多序列比对分析(multiple sequence aligment, MSA)。利用MEGA 4.0软件采用基于距离参数的邻接法(Neighbor-joining, NJ), 并自展分析(Bootstrap)1 000次, 构建各个纲目中模式生物的系统进化树。

1.4 靶基因的预测

以*mmu*-miR-9(小鼠miR-9)为研究对象, 通过TargetScan(http://www.targetscan.org/mmu_50/)、PicTar(http://pictar.mdc-berlin.de/cgi-bin/new_PicTar_mouse.cgi?species=mouse)、microrna.org(http://www.ebi.ac.uk/enright-srv/microcosm/cgi-bin/targets/v5/search.pl)及MicroCosm Targrts(http://www.ebi.ac.uk/enright-srv/microcosm/cgi-bin/targets/v5/search.pl)四个microRNA靶位点分析软件在线分析*mmu*-miR-9的靶基因, 选取四个软件均能预测到的靶基因作为候选靶基因。通过分析靶基因的功能进一步认识*mmu*-miR-9的生物学功能。

1.5 靶基因的功能分析

利用The Gene Ontology(http://www.geneontology.org/)、Kyoto Encyclopedia of Genes and Genomes(http://www.genome.jp/kegg/)和David(http://david.abcc.ncifcrf.gov/)数据库分析靶基因的生物学功能。

2 结果

2.1 *miR-9*基因家族及其在基因组中的定位

在miRBase中搜索后生动物的*mir-9*基因序列, 在47个物种中共搜索到120条*mir-9*基因同源序列。其中, 斑马鱼(*Danio rerio*)的序列数量最多, 共7条, 原鸡(*Gallus gallus*)、大猩猩(*Gorilla gorilla*)、飞蝗(*Locusta migratoria*)等只有一条或者两条序列。值得注意的是, 人类*mir-9*在miRBase中存在*has-mir-9-1*、*has-mir-9-2*和*has-mir-9-3*三条序列, 而同为人类近亲的大猩猩(*Gorilla gorilla*)和黑猩猩(*Pan troglodytes*)在miRBase数据库中分别只有一条和两条*mir-9*序列。利用*has-mir-9*的前体序列作为查询序列, 采用BLAST程序分别对黑猩猩和大猩猩基因组数据库进行搜索, 发现在黑猩猩(*Pan troglodytes*)中还存在一条miRBase中没有收录的*has-mir-9-3*的同源序列, 在大猩猩(*Gorilla gorilla*)中还存在一条miRBase没有收录的*hsa-mir-9-2*的同源序列, mfold折叠证实这些序列都存在miRNA前体典型的二级结构, 说明这些物种中还存在目前尚未发现的新的*mir-9*基因。

以人、黑猩猩、牛、小鼠、斑马鱼、原鸡和黑腹果蝇等7个模式生物*mir-9*前体序列作为查询序列(Query Sequence), 分别对这些生物的基因组DNA进行BLAST搜索, 确定各序列在基因组中的位置(表1)。基因定位显示, 与其它哺乳动物的miRNA主要定位在内含子区域的情况不同, 86%的*mir-9*基因家族成员位于基因间隔区(intergenic region, IGR), 少数位于蛋白编码基因的内含子区域; *mir-9b*、*mir-9c*两个基因位于同一染色体上, 并与*mir-306*、*mir-79*紧密相连聚集成簇; *mir-9a*、*mir-9b*和*mir-9c*主要出现在较低等的模式生物黑腹果蝇中, 而*mir-9-1*至*mir-9-7*则主要存在于哺乳类、鸟类、鱼类等脊椎动物中。*mir-9a*、*mir-9b*和*mir-9c*主要出现在物种进化的早期或低等生物中的现象表明其可能为*mir-9*家族中较原始的基因。

2.2 *miR-9*基因家族序列分析

将*mir-9*基因家族的成熟序列进行多序列比对分析(表2), 发现为第2位到第8位碱基以及第14位到第18位碱基为保守碱基, 说明*miR-9*基因在不同物种间的高度保守性。比对结果显示*miR-9*在其进化过程中可能发生了缺失: 有的物种在第9位缺失了碱基A/U/G, 有的物种在第23位缺失了碱基A; 大多数物种在第12位缺失了碱基U, 第24位缺失了碱基U。第9~11位碱基的相对不保守性可能是因为三个碱基的错配结构更有利于其与靶序列的配对[10]。序列比对发现miR-9与miR-9a的成熟序列相同或只在3'末端相差一个碱基A, 但在所有物种的前体序列的相应位置都具有碱基A, 所以miR-9与miR-9a可能是相同的miRNA, 只是由于3'端的A容易脱落, 不同研究者在实验中得到了3'端相差1个A的两种分子。

作为转录后水平调控元件, miRNA成熟序列5'端的第2-8位碱基对于其功能的正常发挥至关重要, 被称为"种子序列(seed sequence)"[10]。种子序列主要通过与目标mRNA 3'UTR区域的互补配对识别

表 1　miR-9基因家族在七个物种中的基因位置示意图

Table1　Genomic location of miRNA-9 gene family in seven species

物种 Species	miRNA基因名称 Name of miRNA	染色体 Chromosome	基因位置 Gene position
Homo sapiens	*has-mir-9-1*	Chr I	Intron of *Clorf61* gene
	has-mir-9-2	chr5	IGR between *MEF2C* and *TMEM 161B* gene
	has-mir-9-3	chr15	IGR between *Loc100288642* and *Loc254559* gene
Pan troglodytes	*ptr-mir-9-1*	chr1	Intron of *Clorf61* gene
	ptr-mir-9-2	chr5	IGR between *MEF2C* and *TMEM 161B* gene
	Ptr-mir-9-3	chr15	IGR between *PLOG* and *LOC453637* gene
Mouse	*mmu-mir-9-1*	chr3	Intron of *Loc10042277*
	mmu-mir-9-2	chr13	IGR between *MEF2C* and *Loc62324* gene
	mmu-mir-9-3	chr7	IGR between *Plog* and *Rhcg* gene
Gallus gallus	*gga-mir-9-1*	chr21	IGR between *LOC420111* and *LOC427091* gene
	gga-mir-9-2	chrZ	IGR between *LOC769007* and *LOC426919* gene
Bos taurus	*bta-mir-9-1*	chr21	IGR between *Loc533090* and *RHCG* gene
	bta-mir-9-2	chr7	IGR between *TMEM1618* and *MEF2c* gene
Danio rerio	*dre-mir-9-1*	chr16	IGR between *Rhb9* and *Mef 2d* gene
	dre-mir-9-2	chr10	IGR between *LOC564822* and *Mef2ca* gene
	dre-mir-9-3	chr25	IGR between *LOC560613* and *Rhc9* gene
	dre-mir-9-4	chr22	IGR between *tmem161a* and *gata2a* gene
	dre-mir-9-5	chr5	IGR between *mef2cb* and *tmem161b* gene
	dre-mir-9-6	chr7	IGR between *Zgc162132* and *Kif7* gene
	dre-mir-9-7		NOT Found
Drosophila melanogaster	*dre-mir-9a*	chr3L	IGR between *CG93CO* and *Sha1*gene
	dre-mir-9b/9c	chr21	Intron of *Grp* gene，and dre-mir-9b/9c clustering together with mir-306、mir-306s、mir-79

靶基因, 负调控靶基因的表达。在同源性miRNA中, 这段序列通常是保守的, 我们的比对结果也证实了这点。*mir-9*家族前体序列的同源性分析表明, 人与黑猩猩的同源性为100%, 人与恒河猴之间的同源性为85%, 而人与较低等的两栖类(非洲爪蟾)的同源性约为58%, 与果蝇的同源性只有39%。可见亲缘关系近的物种之间核苷酸序列差异小, 亲缘关系远的物种之间核苷酸序列差异大。仔细分析*miR-9*的前体序列发现, 序列差异主要发生在*miR-9*前体中成熟序列以外的区域, miRNA成熟序列的区域由于受到选择压力而高度保守。

2.3 *miR-9*基因家族系统进化树的构建

来源于miRBase中的*miR-9*基因主要分布在哺乳类、鸟类、鱼类、两栖纲和节肢动物门、线虫纲中。采用MEGA 4.0软件, 以这些纲目中的模式物种的*mir-9*序列构建基于距离参数的邻接法(Neighbor-joining, NJ)[11]构建系统发育进化树(图1)。

图1可见, 进化树可分为三支: *mir-9*、*mir-9b*、*mir-9c*主要分布在紫色球海胆、Capitella sp. I(线虫纲)、果蝇(节肢动物门)等无脊椎低等生物中, 聚为第一支。从进化树的分支和距离来看, *mir-9*、*mir-9b*、*mir-9c*出现在进化早期并持续到现在, 由此可以推测*mir-9*、*mir-9b*、*mir-9c*可能是*mir-9*基因家族最早出现的基因形式, 即为祖先基因。而且*mir-9b*和*mir-9c*位于同一分支中, 说明它们亲缘关系近。基因定位分析显示*mir-9b*和 *mir-9c*聚集成簇, 暗示它们在进化过程中紧密相连。第二支主要是分布在节肢动物门的*mir-9a*及部分环节动物门和线虫门的*mir-9*和*mir-9b*。第三分支主要为人、黑猩猩、鱼类、鸟类等脊椎动物的*mir-9*基因家族成员。在这一支中, 除同源性较低、亲缘关系较远的文昌鱼(全索亚门)的*mir-9*单列一小支, 其他物种中的*mir-9-2*、*mir-9-5*、*mir-9-6*属同一亚支, 而*mir-9-1*、*mir-9-3*、*mir-9-4*、*mir-9-7*属与之平行进化的另一亚支。

分析发现, *miR-9b*和 *miR-9c*两个拷贝主要出现在无脊椎动物中, *miR-9-1*、*miR-9-2*、*miR-9-3*、*miR-9-4*、*miR-9-5*、*miR-9-6*和*miR-9-7*主要出现在脊椎动物中, *miR-9*和*miR-9a*在脊椎和无脊椎动物中都有分布。根据上述的分析和比对结果, 可以推测*mir-9c*作为该基因家族最原始的基因存在于线虫纲, 经

表2 mir-9基因家族各成熟序列比对

Table 2 Multiple sequence aligment of mature sequence of mir-9 gene family

miRNA名称 Name of miRNA	成熟序列 Mature sequence	碱基数量 Base numeber	miRNA名称 Name of miRNA	成熟序列 Mature sequence	碱基数量 Base numeber
xtr-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	22	mml-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23
ame-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23	mml-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
bfl-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	22	mml-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
bfl-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	22	lla-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23
bmo-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23	age-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23
lgi-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23	mdo-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
cap-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23	mdo-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
tca-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23	eca-miR-9a-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
lmi-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23	eca-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dpu-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23	cfa-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23
tni-miR-9-4	UCU UUG GUU AU-CUA GCU GUA UGA-	23	cfa-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
tni-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	cfa-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
tni-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	gga-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
tni-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	gga-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
fru-miR-9-4	UCU UUG GUU AU-CUA GCU GUA UGA-	23	tgu-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
fru-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	tgu-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
fru-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	xtr-miR-9a-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
fru-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dya-miR-9a-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-7	UCU UUG GUU AU-CUA GCU GUA UGA--	23	dya-miR-9a-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-6	UCU UUG GUU AU-CUA GCU GUA UGA-	23	aga-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-5	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dan-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-4	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dwi-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	aae-miR-9a-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	aae-miR-9a-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
dre-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	der-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
ssc-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dvi-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
ssc-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dsi-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
rno-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dse-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
rno-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dme-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
rno-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dgr-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
spu-miR-9	UCU UUG GUU AU-CUA GCU GUA UG--	22	dpe-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
bta-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UG--	22	dps-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
bta-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UG--	22	dmo-miR-9a	UCU UUG GUU AU-CUA GCU GUA UGA-	23
oan-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UG--	22	xtr-miR-9a-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23
oan-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UG--	22	xtr-miR-9b	UCU UUG GUU AU-CUA GCU GUA UGA-	23
mmu-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	sme-miR-9a	UCU UUG GUU UU-CUA GCU GUA UGAU	24
sko-miR-9	UCU UUG GUU AU-CUA GCU GUA U---	21	sme-miR-9b	UCU UUG GUU GU-UUA GCU AUA UGA-	25
mmu-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	ame-miR-9b	GCU UUG GUA AU-CUA GCU UUA UGA-	23
mmu-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	bmo-miR-9b	GCU UUG GUA AU-CUA GCU UUA UGA-	23
ppy-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	tca-miR-9b	GCU UUG GUA AU-CUA GCU UUA UGA-	23
ppy-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	aae-miR-9	UCU UUG GU- AUUCUA GCU GUA GA--	22
ppy-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	aga-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22
ptr-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	cqu-miR-9	UCU UUG GU- AUUCUA GCU GUA GA--	22
ptr-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dan-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22
hsa-miR-9-3	UCU UUG GUU AU-CUA GCU GUA UGA-	23	der-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22
hsa-miR-9-2	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dgr-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22
hsa-miR-9-1	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dme-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22

(待续)

(续表2)

miRNA名称 Name of miRNA	成熟序列 Mature sequence	碱基数量 Base numeber	miRNA名称 Name of miRNA	成熟序列 Mature sequence	碱基数量 Base numeber
ggo-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dmo-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22
mne-miR-9	UCU UUG GUU AU-CUA GCU GUA UGA-	23	dme-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
dpe-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22	dmo-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
dps-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22	dpe-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
dse-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22	dps-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
dvi-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22	dse-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
dwi-miR-9c-1	UCU UUG GU- AUUCUA GCU GUA GA--	22	dvi-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
dwi-miR-9c-2	UCU UUG GU- AUUCUA GCU GUA GA--	22	dwi-miR-9b-1	UCU UUG GUG AUUUUA GCU GUA UG--	23
dya-miR-9c	UCU UUG GU- AUUCUA GCU GUA GA--	22	dwi-miR-9b-2	UCU UUG GUG AUUUUA GCU GUA UG--	23
bmo-miR-9c	UCU UUG GU- AUCCUA GCU G-- ----	18	dya-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23
aae-miR-9b	UCU UUG GUG AUUUUA GCU GUA UGC-	23	aga-miR-9b	ACU UUG GUG AUUUUA GCU GUA UG--	23
dan-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23	dgr-miR-9b	UCU UUG GUG AUUUUA GCU CUA UG--	23
der-miR-9b	UCU UUG GUG AUUUUA GCU GUA UG--	23	Clustal consensus	** ****** ******* ****	

阴影区域表示保守碱基, "-"表示碱基缺失。

The shadow region represent highly conserved bases, hyphen (-) indicate base delection.

串联重复后形成了紧密相连的*mir-9b*、*mir-9c*基因簇, 随后经历片段缺失和碱基突变以*mir-9a/9*、*mir-9b*的形式存在于节肢动物门和线虫纲等无脊椎动物中。在由无脊椎动物向脊椎动物进化的过程中*mir-9/9a*基因可能发生片段重复形成了低等脊椎动物(鱼纲)中的*mir-9-1*至*mir-9-7*基因。

2.4 靶基因的预测

大量的研究表明, 一些miRNA在转录后水平调控脑部基因的表达, 而且在神经发育各个阶段都起着重要的调控作用。miRNA在中枢神经系统的表达具短暂性和区域性[12]。如miR-138和miR-124主要在中枢神经系统中表达, miR-126和miR-29主要在星形胶质细胞中表达。据报道miR-9是一个组织特异性的miRNA, 它主要在中后脑交界处(midbrain-hindbrain boundary, MHB)的邻近区域表达[13], 可通过作用于成纤维生长因子(fibroblast growth factor, Fgf)信号通路调控MHB的组织活性, 而且miR-9与神经发生、心脏肥大[14]、脑部正常发育[15,16]等具有密切关系。为进一步了解miR-9参与的调控机制及miRNA作用的新标靶, 本文以小鼠的*mmu*-miR-9基因为研究对象, 采用TargetScan、PicTar、microrna.org和MicroCosm Targrts等四个microRNA靶位点分析软件搜索*mmu*-miR-9的靶基因。利用The Gene Ontology、Kyoto Encyclopedia of Genes and Genomes和David软件分析四个靶基因软件均能预测到的靶基因(表3)的功能, 从而了解miR-9在生物体中可能的作用路径。

由The Gene Ontology和Kyoto Encyclopedia of Genes and Genomes数据库分析可知, miR-9的靶基因多数参与新陈代谢、信号转导、跨膜运输和物质转运等生命活动过程, 与神经细胞、心肌细胞等细胞的生长和分化密切相关。例如: *nidogen 2*所编码的巢蛋白是神经干细胞的标志蛋白, *Erbb2ip*基因所编码的Erbb2相互作用蛋白是神经干细胞正常生长所必须的表皮生长因子受体(epidermal growth factor receptor, EGFR), *Slc31a2*所编码的铜离子转运蛋白对于维持神经系统的正常生理功能非常重要。可见, miR-9在基因表达调控中具有广泛的作用。Chunnian等[17]曾报道miR-9对于神经干细胞的增殖和分化具有调控作用, Krichevsky等[18]在研究胚胎神经干细胞形成时发现以miR-9和miR-124为代表在脑部含量丰富的miRNA, 主要是在神经细胞向神经元的分化过程中诱导产生的, 而且在神经分化过程中阻断它们的功能会提升神经干细胞向星形胶质细胞分化的可能性。这些研究工作也支持本文获得的结果。David数据库的Function annotation chart的数据统计显示*mmu*-miR-9的候选靶基因中, 与磷酸化和乙酰化作用相关的基因数量最多, 暗示着miR-9与表观调控路径相关, 为我们研究mir-9的功能提供一个新的切入点。

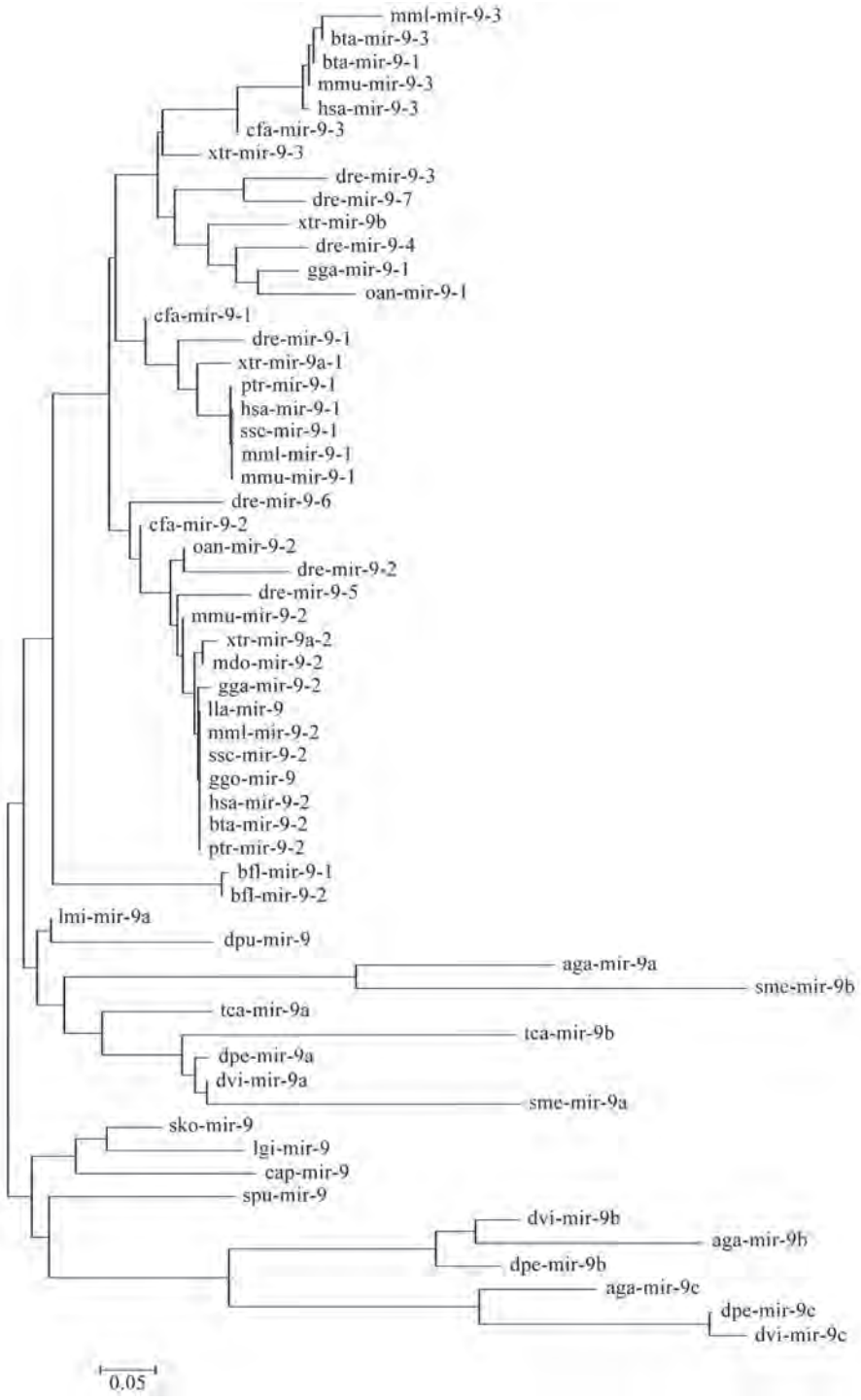

图1 MEGA 4.1软件NJ法构建*mir-9*基因家族系统进化树

缩写含义: bfl: 文昌鱼; xtr: 爪蟾属; gga: 原鸡; cfa: 家犬; mdo: 短尾猞; lla: 绒毛猴属; mml: 恒河猴; ggo: 大猩猩; has: 智人; ptr: 黑猩猩属; oan: 鸭嘴兽; mmu: 小家鼠; bta: 欧洲牛; ssc: 野猪; dre: 斑马鱼; spu: 紫色球海胆; sko: *Saccoglossus kowalevs*ki; dpu: 水蚤属; aga: 冈比亚按蚊; dpe: 果蝇; dvi: 果蝇; lmi: 飞蝗; tca: 拟谷盗属; cap: *Capitella sp. I*; lgi: *Lottia gigantean*; sme: *Schmidtea mediterranea*。

Fig.1 phylogenetic tree of the miR-9 gene family constructed by MEGA 4.1 software based on Neighbor-joining method

Meanings of abbreviations: bfl, *Branchiostoma floridae*; xtr, *Xenopus tropicalis*; gga,*Gallus gallus*; cfa,*Canis familiaris*; mdo, *Monodelphis domestica*; lla, *Lagothrix lagotricha*; mml, *Macaca mulatta*; ggo, *Gorilla gorilla*; has, *Homo sapiens*; ptr, *Pan troglodyte*; oan, *Ornithorhynchus anatinus*; mmu,*Mus musculus*; bta, *Bos Taurus*; ssc, *Sus scrofa*; dre, *Danio rerio*; spu, *Strongylocentrotus purpuratus*; sko, *Saccoglossus kowalevs*ki; dpu, *Daphnia pulex*; aga, *Anopheles gambiae*; dpe, *Drosophila pseudoobscura*; dvi, *Drosophila virilis*; lmi, *Locusta migratoria*; tca, *Tribolium castaneum*; cap, *Capitella sp. I*; lgi, *Lottia gigantean*; sme, *Schmidtea mediterranea*.

表3 四个靶基因分析软件均能预测到的*mmu*-miR-9靶基因

Table 3 Targets of *mmu*-miR-9 predicted by four softwares

NCBI序列号 Acc. No.	基因名称 Name of genes	编码蛋白所处代谢路径及其功能简介 Introduction of gene function
NM_025286	*Slc31a2*, solute carrier family 31, member 2	Solute transport family of 31, also known as copper ion transport protein, central nervous system existence a high concentration of copper ion (Cu^{2+}), and Cu^{2+} play a role in the nerve terminal which can depolarized from the sudden released into the synaptic vesicles contact gap
NM_001002011	*Lmna*, lamin A	Lamin A, the main component of nuclear lamina, to maintain levels in the nucleus plays a supporting role, as the nuclear membrane and chromatin structure provides support. Mainly in hypertrophic cardiomyopathy, arrhythmogenic right ventricular
NM_011960	*Parg*, poly (ADP-ribose) glycohydrolase	cardiomyopathy, cardiomyopathy amplification pathway catalytic activity,Hydrolyzes poly(ADP-ribose) at glycosidic (1"-2') link age of ribose-ribose bond to produce free ADP-ribose., function,Poly(ADP-ribose) synthesized after DNA damage is only present transiently and is rapidly degraded by poly(ADP-ribosegly cohy drolase. Poly(ADP-ribose) metabolism may be required for maintenance of the normal function of neuronal cells., sequence caution, Translated as Met., similarity,Belongs to the poly(ADP-ribose) glycohy drolase family
NM_146151	*Tesk2*, testis-specific kinase 2	Testis-specific protein kinase 2, mainly containing N-terminal kinase do main of serine / threonine protein kinase, mainly expressed in testis and prostate
NM_030706	*Trim2*, tripartite motif protein 2	Tripartite motif protein 2, expression of ihigh levels in the nervous system to maintain its structural plasticity of neurons
NM_019767	*Arpc1a*, actin related protein 2/3 complex,	Actin related protein 2/3 complex and promote microfilament formation of living cells surrounding Mainly involve in FC gamma R-related Shijun effect, adjust the actin cy toskeleton, bacterial metabolic pathways such as epithelial cells infected
NM_199029	*Zfp395*, zinc finger protein 395	Zinc Finger protein 395, and is closely related to embryonic devlopment and cell differentiation is an important function of a class of transcription factor
NM_031397	*Bicc1*, bicaudal C homolog 1 (Drosophila)	Two-tail-C gene homolog 1, play a role in the head and tail structure, the organ development
NM_011843	*Mbc2*, membrane bound C2 domain containing protein	Membrane bound C2 domain containing protein
NM_008695	*Nid2*, nidogen 2	Nidogen 2, one of embryonic intermediate filament protein, which is con sidered a marker of neural stem cells
NM_021563	*Erbb2ip*,Erbb2 interacting protein	Erbb2 interacting proteins for epidermal growth factor receptor (epidermal growth factor receptor, EGFR) Mainly involve in NOD-like receptors primarily signal pathway
NM_016673	*Cntfr*, ciliary neurotrophic factor receptor	Ciliary neurotrophic factor receptors, on nerve cell growth, differentiation has obvious nutritional role Mainly involve in cytokines and cytokine receptor interaction pathway, Jak-STAT signaling pathway
NM_023245	*Palmd*, palmdelphin	Palmdelphin, mainly in the brain, the hippocampus, amygdala
NM_134134	*A630042L21Rik*, RIKEN cDNA A630042L21 gene	High mobility group box domain containing family of complex 3
NM_013890	*Fbxw2*, F-box and WD-40 domain protein 2	F-box and WD-40 domain protein 2
NM_009679	*Ap2m1*, adaptor protein complex AP-2, mu1	Adaptor Protein Complex AP-2, μ1 subunit, mainly catalytic, transport protein, protein binding, lipid binding Mainly involve in endocytosis, Huntington disease and other metabolic pathway

(待续)

(续表3)

NCBI序列号 Acc. No.	基因名称 Name of genes	编码蛋白所处代谢路径及其功能简介 Introduction of gene function
NM_134060	*Slc35b3*, solute carrier family 35, member B3	Family 35 solute transport, membrane transport protein family member, plays an important role in neural stem cell renewal and neurogenesis
NM_026139	*Armcx2*,armadillo repeat containing, X-linked 2	Armadillo repeat containing, X-linked 2, a strong signal in embryonic development stage
NM_011945	*Map3k1*, mitogen activated protein kinase kinase kinase 1	Mitogen-activated protein kinase kinase 1 ,through the MAPK cascade controlled response system involved in cell proliferation and differentiation, when the vitality out of control will lead to tumor Mainly involve in MAPK signal pathway, ubiquitin-mediated protein degradation, RIG-I-like receptor signaling pathway, neurotrophic factor signaling pathway, GnRH signaling path way and other related
NM_015736	*Galnt3*,UDP-N-acetyl-alpha-D-galactosamine	N-acetyl galactose aminotransferase polypeptide, often with the fibro blast growth factor 23 with effect Mainly involve in O-polysaccharide biosynthesis, metabolic pathway
NM_009680	*Ap3b1*, adaptor-related protein complex 3, beta 1 subunit	Adaptor-related protein complex 3, beta 1 subunit, is mainly catalytic, transport protein, protein binding, lipid binding Major metabolic pathways associated with the lysosomal
NM_146164	*Lrch4*, leucine-rich repeats and calponin homology (CH) domain containing 4	Leucine-rich repeats and calponin homology (CH) domain containing 4, embryonic growth and differentiation of specific receptor
NM_146087	*Csnk1a1*, casein kinase 1, alpha 1	Casein kinase 1, in the Wnt pathway plays an important role in Mainly involve in Wnt signaling pathway ,Hedgehog signaling pathway
NM_023670	*Igf2bp3*, insulin-like growth factor 2, binding protein 3	Insulin-like growth factor 2, binding protein 3
NM_178613	*4933433P14Rik*, UPF0279 protein *C14orf1*29 homolog.	RIKEN cDNNA 4933433P14 gene
NM_133195	*Brunol4*, bruno-like 4, RNA binding protein	Bruno-like - Gene 4, RNA binding protein, play a role in alternative splic ing and the target mRNA translation and stability

3 讨论

microRNA作为重要的内源性调控基因, 分布范围广, 参与的生物学过程复杂, 调控靶基因数众多, 对于生物的正常生长发育至关重要, 在动植物的进化过程中扮演着重要的角色。*mir-9*基因家族在原口动物和后口动物中分布广泛, 且高度保守, 说明*mir-9*基因在生物起源和迁移过程中的重要作用。*mir-9*基因家族成熟序列的2-8位碱基即“种子区域”的保守性为100%, 暗示其功能的保守性。从低等向高等进化过程中普遍存在*mir-9*基因家族的调控机制, 说明miRNA的调控机制对物种生存的重要性[19]。

Maller等[20]在通过分析拟南芥基因组, 发现许多miRNAs家族是通过基因组水平上的复制、串联重复或者大片段复制产生的, 接着再发生时空表达的分化。Tanzer等[21]发现*miR-17*基因簇进化历史伴随着串联重复、整个基因簇复制以及个别miRNA的随机缺失, 而且*miR-17*基因家族可能与脊椎动物的早期进化有关。对于*mir-9*基因的进化分析显示, *mir-9*、*mir-9b*、*mir-9c*、*mir-9a*基因主要出现在进化早期, 为祖先基因, 且*mir-9b*和*mir-9c*聚集成簇。Hertel等[22]在后生动物基因组中的miRNA基因家族之间的系统发育分析揭示了串联重复和片段重复是miRNA基因簇进化的主要机制, 也因此产生了miRNA基因簇和miRNA基因家族的多样性分布。由此推测*miR-9b*和*miR-9c*两个基因在同一染色体上所形成的基因簇可能是由串联重复所产生的。*mir-9-1*至*mir-9-7*则主要分布在两栖类及哺乳类等较高等的脊椎动物中, *mir-9*基因家族从无脊椎动物向脊椎动物进化的过程中数量有所增多, 可能发生了串联重复、大片段重复和个别miRNA的缺失。所以对*mir-9*的系统进化分析不仅有利于揭示非编码小分子RNA的起源及其作用机制, 还可以增进对物种进

化的了解。

Kosik等[23]和Landgraf等[24]对于microRNA的研究显示, miRNA在神经系统中含量丰富, 表达具有组织特异性和非对称性, 说明miRNAs在神经系统的特定功能中所扮演的重要角色。Monika等[5]利用原位杂交技术发现miR-124及miR-9在神经系统中表达丰富, miR-9主要在神经前体细胞和神经元中表达, 它是神经发育和成脑中所出现的特异性miRNA[25,26]。本文利用生物信息学方法, 在四个miRNA在线分析软件中搜索miR-9的共同靶基因, 通过GO和KEGG分析显示mir-9基因家族可通过其靶基因作用于多条代谢路径, 且与神经系统正常发育与生长途径具有密切关系。所以对于mir-9基因家族的靶基因的预测不仅有利于了解其可能参与的生物学途径, 还可为以后的实验研究提供理论依据。

参考文献(References)

1 Dennis C. The brave new world of RNA. Nature 2002; 418(6894): 122-4.

2 Ruvkun G. Molecular biology, glimpses of a tiny RNA world. Science 2001; 294(5543): 797-9.

3 Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 2001; 294(5543): 858-62.

4 Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Res 2004; 14(10A): 1902-10.

5 Deo M, Yu JY, Chung KH, Tippens M, Turner DL. Detection of mammalian microRNA expression by *in situ* hybridization with RNA oligonucleotides. Dev Dyn 2006; 235(9): 2538-48.

6 Yoo AS, Staahl BT, Chen L, Crabtree GR. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. Nature 2009; 460(7255): 642-6.

7 Delaloy C, Liu L, Lee JA, Su H, Shen F, Yang GY, *et al*. MicroRNA-9 coordinates proliferation and migration of human embryonic stem cell-derived neural progenitors. Cell Stem Cell 2010; 6(4): 323-35.

8 Ma L, Young J, Prabhala H, Pan E, Mestdagh P, Muth D, *et al*. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nat Cell Biol 2010; 12(3): 247-56.

9 Tavazoie SF, Alarcón C, Oskarsson T, Padua D, Wang Q, Bos PD, *et al*. Endogenous human microRNAs that suppress breast cancer metastasis. Nature 2008; 451(7175): 147-52.

10 Doench JG, Sharp PA. Specificity of microRNA target selection in translationalrepression. Genes Dev 2004; 18(5): 504-11.

11 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4: 406-25.

12 Wang JF, Zhou H, Chen YQ, Luo QJ, Qu LH. Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acids Res 2004; 32(5): 1688-95.

13 Delaloy C, Gao FB. microRNA-9 multitasking near organizing centers. Nat Neurosci 2008; 11: 625-6.

14 Wang K, Long B, Zhou J, Li PF. miR-9 and NFATc3 regulate myocardin in cardiac hypertrophy. J Biol Chem 2010; 285(16): 11903-12.

15 Zhang Z, Chang H, Li Y, Zhang T, Zou J, Zheng X, *et al*. MicroRNAs: potential regulators involved in human anencephaly. Int J Biochem Cell Biol 2010; 42(2): 367-74.

16 Coolen M, Bally-Cuif L. MicroRNAs in brain development and physiology. Curr Opin Neurobiol 2009; 19(5): 461-70.

17 Zhao C, Sun G, Li S, Shi Y. A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination. Nat Struct MolBiol 2009; 16(4): 365-71.

18 Krichevsky AM, Sonntag KC, Isacson O, Kosik KS. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. Stem Cells 2006; 24: 857-64.

19 Massirer KB, Pasquinelli AE. The evolving role of microRNAs in animal gene expression. Bioessays 2006; 28: 449-52.

20 Maher C, Stein L, Ware D. Evolution of Arabidopsis microRNA families through duplication events. Genome Res 2006; 16: 510-9.

21 Tanzer A, Stadler PF. Molecular evolution of a microRNA cluster. J Mol Biol 2004; 339: 327-35.

22 Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, *et al*. The expansionof the metazoan microRNA repertoire. BMC Genomics 2006; 7: 25.

23 Kosik KS. The neuronal microRNA system. Nat Rev Neurosci 2006; 7(12): 911-20.

24 Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, *et al*. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007; 129: 1401-14.

25 Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. Curr Biol 2002; 12: 735-9.

26 Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS. A microRNA array reveals extensive regulation of microRNAs during brain development. RNA 2003; 9(10): 1274-81.

Molecular Evolution of *miRNA-9* Gene Family and Prediction of Their Target Genes

Dan-Dan Luo, Yu-Ping Luo, Juan Peng, Si-Guang Li*

(*College of Life Sciences and Food Engineering*, *Nanchang University*, *Nanchang 330031*, *China*)

Abstract microRNAs are a class of non-coding small RNAs of 21~25 nts in eucaryota, which regulate gene expression at the post-transcriptional level. In this study, we searched *mir-9* genes of metazoa in miRBase, and got a total of 120 sequences in 47 species which indicating the extensive existence of *mir-9* gene in different species. The analysis of gene localization shown 86% of *mir-9* genes locate in the intergenic region(IGR). Multiple sequence alignment of mature sequences of *mir-9* gene family showed that the second to eighth bases and the fourteenth to eighteenth bases are conservative bases. Phylogenetic analysis revealed *mir-9b* and *mir-9c* may be the earliest gene forms of this family, viz. ancestral genes of *mir-9* family. These ancestral genes created *mir-9-1* to *mir-9-7* genes in vertebrate by tandem duplication, larger segment duplication, deletion and mutation of individual bases. Target genes of *mmu-miR-9* are predicted by four miRNA target gene prediction softwares. The result showed that miR-9 was involved in nervous system development, cardiac diseases and transmembrane transport system. These results contribute to further study on the role of microRNA in nervous system development and the mechanism of neural cell growth and differentiation.

Key words *miRNA-9*; evolution; target gene

Received: October 11, 2011 Accepted: February 25, 2011

This work was supported by the National Natural Science Foundation of China (No.30971473, No.31071138)

*Corresponding author. Tel: 86-791-8304099, E-mail: siguangli@163.com