

不同物种 *snR72~snR78* 的比较基因组分析

李 春 罗玉萍*

(南昌大学生命科学与食品工程学院, 南昌 330031)

摘要 采用生物信息学技术对不同物种 *snR72~snR78* (*snR72*、*snR73*、*snR74*、*snR75*、*snR76*、*snR77* 和 *snR78*) snoRNA 基因进行了比较基因组学研究。结果发现, 真菌中 *snR72~snR78* 具有较高的保守性, 它们串联在一起形成基因簇。但在真菌的不同种中, 基因簇中的部分 snoRNA 发生了缺失、位移和重组。动物和植物均存在着真菌 *snR72~snR78* 基因的同源分子。但是, 与真菌 *snR72~snR78* 基因簇不同, 动物和植物中的真菌 *snR72~snR78* 基因的同源分子并不是串联在一起形成基因簇, 而是分散在基因组的不同位点。目前, 在动物中尚未发现 *snR72* 和 *snR75*, 在植物中尚未发现 *snR76*。值得注意的是, 在真菌中, *snR72~snR78* 基因簇的各成员只有一个拷贝, 但在人、动物和植物中, *snR72~snR78* 的一些成员有多个拷贝。人和小鼠的 *snR73* 和 *snR76* 有多个拷贝, 拟南芥 *snR72*、*snR75*、*snR77* 和 *snR78* 有多个拷贝, 水稻 *snR73*、*snR75* 和 *snR77* 有多个拷贝。而且, 一些多拷贝的 snoRNA 基因串联在一起, 表明动植物中的这些多拷贝的基因很可能是在进化过程中由局部复制产生。

关键词 snoRNA; *snR72~snR78* 基因; 基因簇; 比较基因组分析

核仁小分子RNA (small nucleolar RNA, snoRNA) 是一类富集于核仁中, 长约 70~300 nt 的非编码小分子 RNA^[1]。根据 snoRNA 的结构特点, 可以将目前已知的 snoRNA 分为三大类: box C/D snoRNA、box H/ACA snoRNA 和 RNase MRP RNA (mitochondrial RNA processing RNA)。Box C/D 类 snoRNA 一般在其分子 5' 端具有保守的 box C (5'-UGAUGA-3'), 3' 端具有保守的 box D (5'-CUGA-3')。通常情况下, 在 box C 和 box D 之间还有一个 box D' 和 / 或一个 box C' 序列元件。Box D' 和 box C' 中的序列一般仅有一个核苷酸与 box D 和 box C 不同, 这个不同的核苷酸可以位于任何位置。多数 box C/D snoRNA 基因的 5' 和 3' 末端都有 4 nt 以上的反向重复序列, 可以使转录的 snoRNA 前体通过末端碱基配对形成较为稳定的短茎结构。大部分 box C/D snoRNA 的 box D 或 D' 的上游都包含一段长 10~21 nt 的片段, 称为反义序列, 能与 rRNA 或 snRNA 中一段靶序列互补, 直接参与指导 rRNA 的 2'-O- 核糖甲基化修饰^[2]。

迄今, 人们发现 snoRNA 的基因组织主要有三种类型, 即独立转录的 snoRNA 基因组织、内含子编码的 snoRNA 基因组织和 snoRNA 基因簇。不同生物体的 snoRNA 具有不同的基因组织, 几乎所有的脊椎动物的 snoRNA 由内含子编码, 而酿酒酵母的 snoRNA 则大多由独立基因编码, 少量由内含子编码

或以成簇的形式存在^[2-4]。植物 snoRNA 的基因组织相当复杂, 大多数植物 snoRNA 以基因簇的形式存在, 以多顺反子的方式表达^[5-8], 而且内含子中成簇的 snoRNA 基因在水稻中相当普遍^[7]。

snoRNA 的缺失与人类某些疾病密切相关, 印记表达的 snoRNA 基因簇 MBII-85 的缺失会导致一种复杂的神经发育障碍疾病——帕 - 魏二氏综合征 (低肌张力 - 低智能 - 性发育低下 - 肥胖综合征)^[9]。因此, 采用比较基因组学的方法研究 snoRNA 基因簇的缺失、插入和进化具有重要意义。目前在酿酒酵母中发现了 5 个 snoRNA 基因簇, 其中最复杂的基因簇由 7 个 box C/D snoRNA 基因 (*snR72~snR73~snR74~snR75~snR76~snR77~snR78*) 组成^[10], 本文将该基因簇简称为 *snR72~snR78* 基因簇。研究表明, 这个基因簇首先在上游启动子的作用下转录出多顺反子 snoRNA 前体, 然后在核酸内切酶 RNase III 的作用下使每个 snoRNA 分离, 最后 snoRNA 经核酸外切酶 Rat1p, Xrn1p (5' → 3' 活性) 和 Rps6p (3' → 5' 活性) 的作用加工成熟^[10]。

snR72~snR78 基因簇的成员广泛分布于真菌、

收稿日期: 2009-08-07 接受日期: 2009-11-24

国家自然科学基金(No.30660042)和江西省自然科学基金(No.0630136)资助项目

* 通讯作者。Tel: 0791-8304938, E-mail: luoyuping@163.com

植物和动物中。但是,在进化过程中,该基因簇的成员发生了缺失、插入、移位和重组等分子事件,在不同的物种中表现出不同的基因组织形式,因此是研究非编码 RNA 基因簇进化的理想分子。本文采用生物信息学技术对不同物种中酵母 *snR72~snR78* snoRNA 基因簇的同源分子进行研究,系统分析在进化进程中该基因簇中各成员的缺失、移位和重组规律,以期阐明小分子 RNA 的分子进化及小分子 RNA 基因簇在基因组进化中的作用。

1 材料与方法

1.1 人、水稻、拟南芥和酿酒酵母 *snR72~snR78* 序列的获得

从三个 snoRNA 数据库(<http://lowelab.ucsc.edu/snoRNAdb/>; <http://www-snoRNA.biotoul.fr/>; http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/)中获得人、水稻、拟南芥和酿酒酵母的 *snR72~snR78* 序列。然后用已获得的序列对人、水稻、拟南芥和酿酒酵母的核酸数据库进行同源性搜索,分析这些生物的基因组中是否还存在 snoRNA 数据库中尚未报道的 *snR72~78* 同源序列。

1.2 小鼠和其它真菌中 *snR72~snR78* 的同源性搜寻

分别用人和酿酒酵母的 *snR72~snR78* 序列对小鼠和其它真菌的核酸数据库进行同源性搜寻。本文使用了三个数据库,分别是 NCBI 数据库(<http://www.ncbi.nlm.nih.gov/blast>)、ENSEMBL 数据库(<http://www.ensembl.org/index.html>)以及 UCSC 数据库(<http://genome.ucsc.edu/cgi-bin/hgBlat>)。采用 BlastN 程序^[11]对 NCBI 和 ENSEMBL 数据库进行搜寻,采用 BLAT 程序^[12]对 UCSC 数据库进行搜寻。

1.3 不同物种 *snR72~snR78* 序列的比对分析

采用 Clustal W 程序^[13]对不同物种的 *snR72~snR78* 进行多序列比对(multiple sequence alignment, MSA)分析。

2 结果

2.1 真菌中的 *snR72~78* snoRNA 基因簇

通过同源性搜索的方法,分别在 *Saccharomyces paradoxus*、*Saccharomyces mikatae*、*Saccharomyces kudriavzevii*、*Saccharomyces bayanus*、*Candida glabrata*、*Saccharomyces castellii*、*Kluyveromyces lactis*、*Kluyveromyces waltii* 和 *Eremothecium gossypii* 中发现了由 7 个 box C/D snoRNA 基因(*snR72~snR78*)

组成的基因簇,而在 *Pichia guilliermondii* 中只有由 6 个 box C/D snoRNA 基因(*snR73~snR74~snR75~snR76~snR77~snR78*)组成的基因簇,在粟酒裂殖酵母中由 5 个 box C/D snoRNA 基因(*snR73~snR74~snR75~snR77~snR78*)组成的基因簇(图 1)。有趣的是,虽然粗糙脉胞菌(*Neurospora crassa*)也是由 6 个 box C/D snoRNA 基因(*snR73~snR74~snR75~snR76~snR77~snR78*)组成基因簇,但其中一个 snoRNA 具有 *snR72* 和 *snR78* 的分子特征(即具有 *snR72* 和 *snR78* 分子的反义序列)(图 1)。以上结果表明,虽然该 snoRNA 基因簇在真菌中具有较高的保守性,但在进化过程中,基因簇中的一些 snoRNA 发生了缺失、移位和重组。

2.2 真菌 *snR72~snR78* snoRNA 基因簇在动物中的同源分子

在人和小鼠中存在着真菌 *snR72~snR78* 基因簇的同源分子。值得注意的是,与真菌 *snR72~snR78* 基因簇不同,动物中与该基因簇中各成员同源的 snoRNA 散布在基因组的不同位置,而不是串联在一起形成基因簇(图 2)。有趣的是, *snR76* 在真菌中只

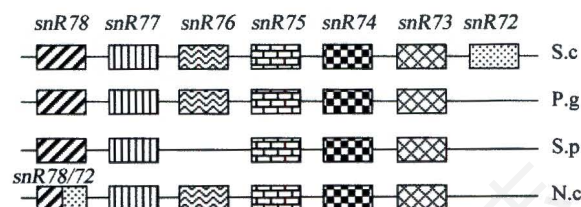


Fig.1 *snR72~snR78* snoRNA gene clusters in different fungus species

S.c: *Saccharomyces cerevisiae*; P.g: *Pichia guilliermondii*; S.p: *Schizosaccharomyces pombe*; N.c: *Neurospora crassa*.

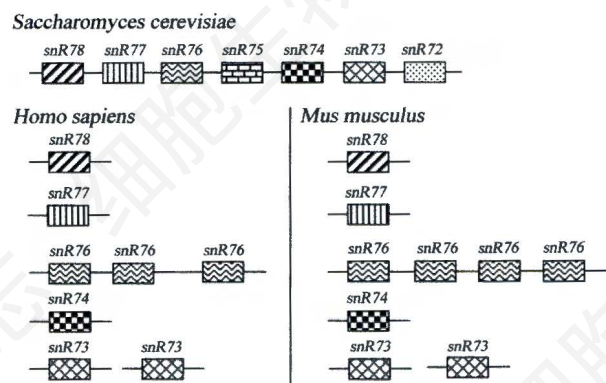


Fig.2 The homologues of fungi *snR72~snR78* snoRNA gene cluster in animal

snoRNA	Spe.	Chr	Start	Sequence alignment
snR76				
S.c	XIII	297727	GGAAATGAGGATCCACAGCACTGGGCACTGAGGGGGAAACAAATATCCTTTCAAATGAG-----TGACAATGCTTTA--GCGGGCCT-AGAAAACCCGCGCAACTGATCCA-----	
P.g	cont1.61*	140765	-ACAAATGAGGAAATTATTGCACCTGGGCACTGAAAG-----CCATGAG-----TGATAACGCATTAAAGGTGAGCCTCAGAAACTCACCGTGAATGAGTGA-----	
N.c	VI	244682	TTCTATGATGAA--ACGGCACTGGGCTCCGA-----TTTTTCCACATGAGCAGCATATGCCATTTTTCGCG--ACAGGCCT-AGAAAACCTCGTCCACTGATCAGC-----	
H.s	19	23570975	CTCCATGATGTC--CAGCACTGGGCTCCGACTG-----CCACTGAGGAC-----ACGGTGCCCC--CCGGGACCTTTGACACCCGGGG-GTCTGAGGGGC-----	
H.s	19	23570565	CCCCGATGATGTC--CAGCACTGGGCTCTGACTG-----CCCCTGAGGAC-----ACGGTGCACC--CCGGGACCTTTGACATCCGGGG-TTCTGAGGGGC-----	
H.s	19	23573861	TCCCATGATGTC--CAGCACTGGGCTCTGATCA-----CCCTGAGGAC-----ACAGTGCACC--CCAGGACCTTTGACACCTGGGG-GTCTGAGGGGC-----	
M.m	7	5149037	-----TGATGTT--CAGCACTGGGCTCTGACCA-----TC-CTGAGGAC-----ATGGTGCCCC--CAGGGACCTTTGACACCTGGG-GTCTGAGGGGCCCTGG-----	
M.m	7	5149315	CCCCTGATGTC--CAGCACTGGGCTCTGACCA-----TC-CTGAGGAC-----ATGGTGCCCC--CAGGGACCTTTGACACCTGGG-GTCTGAGGGGTCCTG-----	
M.m	7	5147275	-----TGATGTT--AGCACTGGGCTCCAAACA-----CCCTTGAAGGAC-----ACAGTGCCCTT--CCAGGACCTTTGACACCTGGAG-ATCTGAGG-----	
M.m	7	5149630	CCCCATGATGTTAT-ATAGCACTGGGCTCTGACCA-----CC-ATGGTGAT-----ATGTTGTGCTT--CCAGGACCTTTGACATCCAGGA-GTCTGAGGGG-----	
snR77				
S.c	XIII	297506	TCATATGATGATTATA--ACAAAACAAAAGCTGGAATTACTGGCTGACGA--GTAT-----ATGTTGATACG----TTTTTGCACTATC-CTGACCTA----	
P.g	cont1.61*	140629	TTAGATGATGACT-TA--ACAAAACAAA--GCTGGAATTACTGGCTGATCT--CTAT-----G-ATTGACAAG----TTTTTGCACTATTTCTGATTTACCA-----	
S.p	I	526938	GGAAATGATGAAAA--ACTTTT--TTTGGAAATTACCGGCTAGGAATAGTTTACTCTCTATTCGCTGAAGAATAAATGTGCTGCAATTATGCTGATCCAT--	
N.c	VI	244445	TTTCTGATGATCAACTTTGCTTTTGCA--AGTTGGAATTACCGATTGA-----TAC-----CCAATGAAAA--A-TAAAAATGGCTGTTTCTGATTGAA--	
A.t	2	7520943	-CCGATGATGATTAT--TGCGAAA--ATATGGAATTACCGTCTGAGTTAATTTCT-----TTGATGTTAAA--AAC--GCTGGCTAACTGAGG-----	
A.t	2	2184645	ACCGATGATGATTAT--TGCTAAA--CTATGGAATTACCGTCTGAGTTTA--TTCT-----TTGACGAGAAA--AAC--GCTGGCTTACTGAGGT-----	
O.s	3	30968948	GCCTGTGATGATTA-----GCATAT--TGTTGGAATTACCGTCCGA-----TTAT-----CAAATGACGAC--ATTGTCTGGCTTCCTGAGGCTT--	
O.s	3	30971046	GCCTGTGATGATTA-----GCATAT--TGTTGGAATTACCGTCCGA-----TTAT-----CAAATGACGAC--ATTGTCTGGCTTCCTGAGGCTT--	
O.s	3	30973146	GCCTGTGATGATTA-----GCATAT--TGTTGGAATTACCGTCCGA-----TTAT-----CAAATGACGAC--ATTGTCTGGCTTCCTGAGGCTT--	
O.s	3	30966852	GCCTGTGATGATTA-----GCATAT--TGTTGGAATTACCGTCCGA-----TTAT-----CAAATGACGAC--ATTGTCTGGCTTCCTGAGGCTT--	
O.s	3	30975265	GCCTGTGATGATTA-----GCATAT--TGTTGGAATTACCGTCCGA-----TTAT-----CAAATGACGAC--ATTGTCTGGCTTCCTGAGGCTT--	
O.s	2	26282632	GCCTATGATGATTA-----GCATAT--TGTTGGAATTACCGTCCGA-----TTAT-----CAAATGACGAC--ATTGTCTGGCTTCCTGAGGCTT--	
H.s	17	15941889	--AAATGATGAATCA--CCCAAAATA--GCTGGAATTACCGGCAGATTGT--GTAG-----TGGTGAA--CCTATGG--TTTTCTGAG-----	
M.s	11	27922895	--AAATGATGATATCA--CCTAAAATA--GCTGGAATTACCGGCAGATTGT--GTAG-----TGGTGAG--CCTATGG--TTTTCTGAG-----	
snR78				
S.c	XIII	297281	-----TTGATGACCAAAATAAATTTTCAAACTAGAGTTTCTGAA-----TCTTTTGTGATTAGAACGTT-TATTCTTATTT--CTGAGGTT-----	
P.g	cont1.61*	140461	---CAATGAGGATGATC-CAAAAATTTTCAAACTAGAGTTTCTGAC-----TACGGAATGATCACAGCGCATATGCATGGTTAACTGAGTTGCT---	
S.p	I	526726	-ATCTTTGATGACCATTTTTTAAAT-ACAACTAGAGTTTCTGAT-----TAATTTATGATTTCAA-----ATTCTTG--CTGAGTTTAGT--	
N.c	VI	244084	ATGTCTGATGACACAACCATAAAC-GCAACTAGAGTCTTTGAT-----CCTACGGGAATGCAACTAGATATGTACGTGAGTCTGATGACAA--	
A.t	3	7682361	TTATGGTATGATGAT--AACGAATATTTCTGGACTAGAGTTTCTGATCTGGG--CTCTTACCCAGAAATTTGAAGAAACAACCTTTGGCTGTCTGAGTAAAT---	
A.t	4	8459710	ATATGGTATGATGAT--TACAAGAGTATTGTGGACTAGAGTTTCTGATCTGGGATTCTTCTCCAGAAAGTTGAAGAT-TAACCCTTGGCTGTCTGAGTATT---	
A.t	4	8713335	ATATGGTATGATGAT--TACAAGAGTATTGTGGACTAGAGTTTCTGATCTGGGATTCTTCTCCAGAAAGTTGAAGAT-TAACCCTTGGCTGTCTGAGTATT---	
H.s	6	22663104	-GGGAATGATGAT-----TTTCACAGACTAGAGTCTCCGATG-----CTGGTCATGATGTCAAAAC-TAAGTT-----CTGAG-----	
M.m	17	21281749	TGAGAGTATGATGAT-----TTTCACAGACTAGAGTCTCTGACG-----CTGTCCTTGATGTGACGCTA-TAAAT-----CTGACTCATTGG	

Boxes C, D', C' and D are shown in open boxes, antisense sequences are in underlines. Spe.: species; Chr.: chromosome; Start: chromosome start position; S.c: *Saccharomyces cerevisiae*; P.g: *Pichia guilliermondii*; S.p: *Schizosaccharomyces pombe*; N.c: *Neurospora crassa*; H.s: *Homo sapiens*; M.m: *Mus musculus*; A.t: *Arabidopsis thaliana*; O.s: *Oryza sativa*; cont1.61*: *Pichia guilliermondii* ATCC 6260 cont1.61.

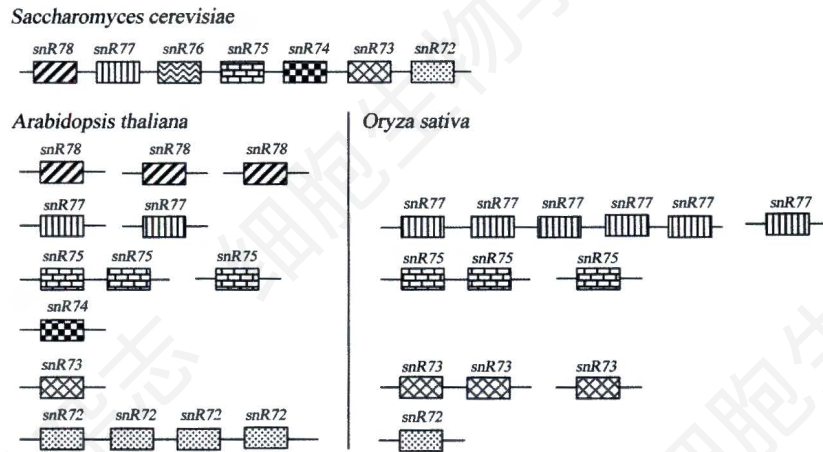


Fig.3 The homologues of fungi *snR72~snR78* snoRNA gene cluster in plant

有一个拷贝,而在人类有三个拷贝,在小鼠中有四个拷贝。*snR76* 的多个拷贝串联在一起形成基因簇(图2),这一组织形式表明动物中这些多拷贝的 snoRNA 基因很可能是在进化过程中由局部串联重复产生。另外,动物中未发现 *snR72* 和 *snR75* 的同源分子,表明在动物的进化过程中这些分子发生了缺失。

2.3 真菌 *snR72~snR78* snoRNA 基因簇在植物中的同源分子

在水稻和拟南芥中存在着真菌 *snR72~snR78* snoRNA 基因簇的同源分子。和动物一样,与真菌 *snR72~snR78* 基因簇中各成员同源的植物 box C/D snoRNA 基因不是串联在一起形成基因簇,而是分散在基因组的不同位点。与动物不同的是,动物中 *snR76* 有多个拷贝,且成簇排列,而在植物中尚未发现 *snR76* snoRNA。但是,在拟南芥中, *snR72*、*snR75*、*snR77* 和 *snR78* 有多个拷贝,在水稻中, *snR73*、*snR75* 和 *snR77* 有多个拷贝。而且植物中大多数多拷贝的 snoRNA 串联在一起(图3)。实际上,水稻中所有这些 snoRNA 基因又和其它 snoRNA 基因串联在一起,形成更大的基因簇。

2.4 *snR72~snR78* snoRNA 的序列分析

snR72~snR78 snoRNA 的序列分析显示,真菌、动物和植物 *snR72*、*snR73*、*snR74*、*snR75*、*snR76*、*snR77* 和 *snR78* 的 box C/D 基序和反义序列等功能序列在进化过程中由于受到选择压力而高度保守,其余部分则变化很大(表1)。不同的生物中,这些 box C/D snoRNA 序列长度也不一致,如酿酒酵母的 *snR73* snoRNA 序列长度为 109 nt,而小鼠的 *snR73* snoRNA 的长度只有 81 nt。

3 讨论

snoRNA 是从古菌到哺乳动物中均广泛存在的非编码小分子 RNA,在 rRNA 和 snRNA 的转录后加工修饰过程中起着重要的作用。古菌、真细菌和真核生物三者都起源于共同的祖先 - 原始细胞,现已发现古菌和真细菌之间存在越来越多的差异,而古菌在有些方面比真细菌更接近真核生物^[14]。近年来,人们在古菌中发现了 snoRNA 及其相关蛋白的类似物^[15],说明 snoRNA 及其结合蛋白具有古老的进化起源,它们在古菌与真核生物分化之前就已存在,距今已有 20 多亿年的进化史^[16]。

对 snoRNA 基因的系统进化分析将有助于进一步了解其起源和进化机制。Zemann 等^[17]在研究线虫 snoRNA 进化时指出, snoRNA 基因的进化主要是通过反式重复(*trans-duplication*)和顺式重复(*cis-duplication*)的策略来完成的。Luo 等^[18]在研究人的 box H/ACA snoRNA 时发现, snoRNA 可以通过返座子机制而复制。此外, snoRNA 与其 rRNA 靶序列存在共进化的现象。本研究表明, *snR72~snR78* snoRNA 基因在从真菌向高等动植物的进化过程中发生了基因的复制、缺失、移位和重组(图1~图3),其基因复制主要表现为基因的串联重复和局部重复。这些发现对于进一步揭示非编码 RNA 基因簇的进化规律具有重要意义。

参考文献(References)

- [1] Maxwell ES, Fournier M J. The small nucleolar RNAs, *Annu Rev Biochem*, 1995, 64: 897-934
- [2] Lowe TM, Eddy SR. A computational screen for methylation guide snoRNAs in yeast, *Science*, 1999, 283(5405): 1168-1171

- [3] Samarsky DA, Fournier MJ, Singer RH, *et al.* The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization, *EMBO J*, 1998, 17(13): 3747-3757
- [4] Qu LH, Henry Y, Nicoloso M, *et al.* U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28S rRNA, *Nucleic Acids Res*, 1995, 23(14): 2669-2676
- [5] Qu LH, Meng Q, Zhou H, *et al.* Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*, *Nucleic Acids Res*, 2001, 29(7): 1623-1630
- [6] Brown JW, Clark GP, Leader DJ, *et al.* Multiple snoRNA gene clusters from *Arabidopsis*, *RNA*, 2001, 7(12): 1817-1832
- [7] Chen CL, Liang D, Zhou H, *et al.* The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*, *Nucleic Acids Res*, 2003, 31(10): 2601-2613
- [8] Leader DJ, Sanders JF, Waugh R, *et al.* Molecular characterization of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts, *Nucleic Acids Res*, 1994, 22(24): 5196-5203
- [9] Ding F, Li HH, Zhang S, *et al.* SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice, *PLoS One*, 2008, 3(3): e1709
- [10] Qu LH, Henras A, Lu YJ, *et al.* Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast, *Mol Cell Biol*, 1999, 19(2): 1144-1158
- [11] Altschul SF, Gish W, Miller W, *et al.* Basic local alignment search tool, *J Mol Biol*, 1990, 215(3): 403-410
- [12] Kent WJ. BLAT — the BLAST-like alignment tool, *Genome Res*, 2002, 12(4): 656-664
- [13] Thompson JD, Higgins DC, Gibson TJ. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res*, 1994, 22(22): 4673-4680
- [14] Olsen GJ, Woese CR. Archaeal genomics: an overview, *Cell*, 1997, 89(7): 991-994
- [15] Gaspin C, Cavaille J, Erauso G, *et al.* Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes, *J Mol Biol*, 2000, 297(4): 895-906
- [16] Terns MP, Terns RM. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin, *Gene Expr*, 2002, 10(1-2): 17-39
- [17] Zemann A, Bekke A O-D, Kiefmann M, *et al.* Evolution of small nucleolar RNAs in nematodes, *Nucleic Acids Res*, 2006, 34(9): 2676-2685
- [18] Luo Y, Li S. Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs, *Nucleic Acids Res*, 2007, 35(2): 559-571

Genome-wide Comparative Analysis of *snR72–snR78* Genes from Different Species

Chun Li, Yu-Ping Luo*

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

Abstract Genome-wide comparative analysis of *snR72–snR78* (*snR72*, *snR73*, *snR74*, *snR75*, *snR76*, *snR77* and *snR78*) genes from different species was performed with bioinformatics technology. Results shown that the *snR72–snR78* genes in fungi are well conserved and linked as a cluster in head-to-tail fashion, while some genes in this cluster have undergone deletion, translocation and recombination during the evolution of different fungus species from primitive progenitor fungi. Animals and plants have homologues of *snR72–snR78* of fungi, but in contrast with fungi in which *snR72–snR78* snoRNA genes are arranged in a cluster, *snR72–snR78* genes in animals and plants are scattered at different sites of the genomes. Up to now, *snR72* and *snR75* do not been found in animals and *snR76* dose not been found in plants. Remarkably, each member of *snR72–snR87* genes in fungi have only one copy, while some of them in human, animals and plants have multiple copies. *SnR76* gene has multiple copies in both human and mouse. *SnR72*, *snR75*, *snR77* and *snR78* have multiple copies in *Arabidopsis* and *snR73*, *snR75* and *snR77* have multiple copies in rice. Moreover, multiple copies of some snoRNA genes closely link to form snoRNA gene clusters, suggesting extensive local tandem duplications have taken place in animal and plant evolution.

Key words snoRNA; *snR72–snR78* genes; gene cluster; genome-wide comparative analysis

Received: August 7, 2009

Accepted: November 24, 2009

This work was supported by the National Natural Sciences Foundation of China (No.30660042) and the Natural Science Foundation of Jiangxi Province (No.0630136)

*Corresponding author. Tel: 86-791-8304938, E-mail: luoyuping@163.com