

U83 Box C/D snoRNA 构建果蝇科系统发生树

徐玲^{1,2} 罗玉萍¹ 周冬根¹ 李思光^{1*}¹南昌大学生命科学学院, 南昌 330031; ²江西科技师范学院生命科学学院, 南昌 330013)

摘要 利用已报道的黑腹果蝇 U83 基因搜索果蝇基因数据库, 鉴定了 10 种新的果蝇科 U83 同源基因, 它们均位于相应物种核蛋白基因 rpl3 的内含子中。以冈比亚按蚊为外类群, 对 11 种果蝇的 U83 核苷酸序列作进化关系分析, 用邻接法重建了系统发生树, 结果与传统方法构建的系统发生树相比, 能反映果蝇科的大致进化关系, 但还存在部分差别。为增加序列信息, 把序列长度拓展至整个 U83 所在的内含子, 同法构建系统发生树, 结果与传统系统发生树几乎完全一致。该研究是用 box C/D snoRNA 基因序列构建系统发生树的首次尝试, 实验结果证明 U83 可以很好地用于构建果蝇科内各物种的种系发生树。

关键词 U83 snoRNA; 内含子; 系统发生树; 果蝇科

核仁小分子 RNA (small nucleolar RNA, snoRNA) 是一类非编码小分子 RNA。除 MRP RNA (mitochondrial RNA processing RNA), 它可分为 box C/D snoRNA 和 box H/ACA snoRNA 两大类, 作用于 rRNA 的剪切、加工和成熟, 指导 rRNA、snRNA、tRNA 碱基修饰和 mRNA 的编辑, 有的还具有维持端粒酶稳定性等功能^[1,2]。

Box C/D snoRNA 的序列特征非常明显。序列的两端具有异常保守的 C box (ugauga) 和 D box (cuga); 链的中部通常还存在另一组保守性较弱的 C、D box, 被称为 C'、D' box; 在 D (或 D') box 的上游通常有一段能和靶标 RNA (如 rRNA 或 snRNA) 互补的序列, 称做反义序列 (antisense element, AE), 通过碱基互补配对, 指导相关的酶对特异位点进行甲基化修饰。在 box C/D snoRNA 序列中, 只有 C、D box (和 C'、D' box) 以及 AE 序列受选择压力而保守, 其他的核苷酸没有选择压力而高度可变^[3,4]。

U83 是 box C/D snoRNA 中的一种, 它在从昆虫到哺乳动物中都广泛存在, 并保守地位于核蛋白 rpl3 基因的内含子当中^[5]。果蝇科的多个物种 DNA 序列在果蝇序列数据库都有公布。本研究中, 我们充分利用以上公共资源, 在 10 种果蝇序列数据库中找到了 U83 的基因, 对它们的序列特征进行了分析, 并分别利用这些序列和整个 U83 所在的内含子序列重建了系统发育树, 与果蝇科物种树进行对比, 判断新建进化树的正确程度, 探讨 box C/D snoRNA 基因在群体系统演化研究中的价值。

1 材料与方法

1.1 使用的数据库

FlyBase 果蝇数据库 (<http://flybase.net>)。

1.2 U83 snoRNA 的搜索

将黑腹果蝇 DmU83 基因序列作为比对序列, 在 FlyBase 果蝇数据库 (<http://flybase.net>) 中用 BLASTN 程序 (<http://flybase.net/blast/>) 搜索 10 种果蝇 *Drosophila simulans*, *Drosophila sechellia*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila ananassae*, *Drosophila pseudoobscura*, *Drosophila persimilis*, *Drosophila virilis*, *Drosophila mojavensis*, *Drosophila grimshawi* 中和冈比亚按蚊 *Anopheles gambiae* 中的 U83 序列。用 PC gene6.0 软件包对所获得的数据做进一步序列鉴定和分析。

1.3 U83 所在内含子的搜索

分析所有 U83 基因两侧序列, 寻找对应于 DmU83 所在内含子两端剪接位点的序列。

1.4 系统发生树的构建

对 U83 基因和 U83 所在内含子序列分别采用 MEGA3.1 软件^[6]分析序列信息, Clustal X 软件^[7]比对序列, PAUP4.0 软件构建系统进化树。

1.4.1 用 U83 序列构建果蝇科系统发生树 用 clustalX1.83 比对所有 12 种 U83 序列 (以冈比亚按蚊为外类群), MEGA3.1 软件计算碱基组成, 转换颠换

收稿日期: 2007-04-04 接受日期: 2007-06-15

国家自然科学基金资助项目 (No.30660042)

* 通讯作者。Tel: 0791-8304099, E-mail: siguangli@163.com

比等。构建的分子系统树采用 PAUP4.0 软件的邻接法(neighbor joining)。U83 基因在碱基组成上表现出 A+T 偏向性,故遗传距离采用 P distance 方法计算。自展法(Bootstrap)检验系统树,自展数据集为 1 000 次。

1.4.2 用 U83 所在内含子序列构建果蝇科系统发生树 12 种 U83 所在内含子序列用 Clustal X 比对后,手工适当校正以使序列排列更加合理。MEGA3.1 软件计算碱基组成,转换颠换比等。用 PAUP4.0 软件邻接法构建系统发生树。

2 结果

2.1 新 U83 序列及其侧翼序列的获得

通过对现有果蝇数据库 *Drosophila simulans*, *Drosophila sechellia*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila ananassae*, *Drosophila pseudoobscura*, *Drosophila persimilis*, *Drosophila mojavensis*, *Drosophila virilis*, *Drosophila grimshawi* 10 个物种的序列进行计算机方法搜索,找到了以下 10 种新的 U83 box C/D snoRNA 基因和用于构建外类群的冈比亚按蚊 U83 box C/D snoRNA 基因(图 1)。

2.2 序列特征

所有 10 种果蝇 U83 基因都具有保守的 C、D 盒;下游 AE 序列也十分保守;两端有反向互补序列形成碱基配对的茎状结构,但序列差异较大,即二级结构保守,一级结构不保守。分析 U83 侧翼序列发现:新鉴定的 10 种果蝇 U83 基因都位于相应物种的核蛋白 rpl3 基因的内含子中,且所在的内含子都遵循 GT/AG 剪接方式,剪接位点序列保守。U83 序列核苷酸数目相差不大,序列长度在 80~89 nt 之间;U83 所在的内含子序列核苷酸数目相差较大,长度在 244~360 nt 之间不等。

U83 序列碱基组成: T 33.2%、C 20.7%、A 24.7%、G 21.3%, A+T 57.9% 高于 C+G 42%。转换与颠换的比值 0.8。对 U83 所在的整个内含子序列进行分析后,得到的统计学特征为: T 32.1%、C 21.5%、A 26.5%、G 20.0%, A+T 58.6% 高于 C+G 41.5%。转换与颠换的比值为 0.7。

有文献指出^[8],当转换与颠换的比值小于 2 时,说明该基因序列的突变已达到饱和状态,进化受饱和效应的影响,有进化噪音,重建系统发生树时如不进行特别加权就会得出错误信息。故分别根据 U83 基因

<i>D. melanogaster</i>	GTCTGGATGATGAAACGATGGTATCTGAATGAGCAAA-TTATTGTTGAGAAGCCT----T
<i>D. simulans</i>	GTCTGGATGATGAAACGATGGTATCTGAATGAGCAAA-TTATTGTTGAGAAGCCT----T
<i>D. sechellia</i>	GTCTGGATGATGAAACGATGGTATCTGAATGAGCAAA-TTATTGTTGAGAAGCCT----T
<i>D. yakuba</i>	GTCTGGATGATGAAAAATGGTATCTGAATGAGCAAA-TTATTGTTGAAA--CCT--T
<i>D. erecta</i>	GTCTGGATGATGAAACGATGGTATCTGAATGAGCAAA-TTATTGTTGAGAAGCCT----T
<i>D. ananassae</i>	GTTTGGATGATGATACGATGGTATCTGAATGAGCAAA-TTCCTGTTGAAAAGCCT----T
<i>D. persimilis</i>	ACCTGAATGATGATACGATGGTATCTGAATGAGCAAAATCTTGTGAAAAACCT----T
<i>D. pseudoobscura</i>	ACCTGAATGATGATACGATGGTATCTGAATGAGCAAAATCTTGTGAAAAACCT----T
<i>D. virilis</i>	TGCTGGATGATGATGAGATGGTATCTGAATGACC-----TTGTGCAAAAACCT----A
<i>D. grimshawi</i>	TGCTGAATGATGATAAGATGGTATCTGAATATAT-----GTGTGCAAAAAGCCT----T
<i>D. mojavensis</i>	TTCTGTATGATGAGAATTGGTATCTGAATGACC-----CTGTGCAAAAATCTTG--T
<i>A. gambiae</i>	GCTCTTATGATGTAATTCCGCCCTAGTGGCAAACCTGTCTGAATATTCATGTACGAAGAAT
	***** * * * * *
<i>D. melanogaster</i>	TTACCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. simulans</i>	TTACCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. sechellia</i>	TTACCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. yakuba</i>	TGAGCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. erecta</i>	CTACCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. ananassae</i>	CTACCAT--CTGCTGCCTTCCTTCGATCAG
<i>D. persimilis</i>	CTACCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. pseudoobscura</i>	CTACCAT--CTGCTGCCTTCCTTCGATCGG
<i>D. virilis</i>	CTACCATTTCTGCTGCCTTCCTTCGATCGC
<i>D. grimshawi</i>	CTACCATCACTGCTGCCTTCCTTCGATATC
<i>D. mojavensis</i>	ATACCAT--ACTGTTGCCTTCCTTCGAAAC
<i>A. gambiae</i>	CTGACAT--ATGCTGCCTTCCTTCGATCAG
	*** ** ***** **

图 1 果蝇科 U83 box C/D snoRNA 基因及冈比亚按蚊 U83 基因序列

星号表示保守序列,方框内为 box C/D snoRNA 的特征序列,依次分别为 C 盒、D 盒。

和 U83 所在内含子序列的转换与颠换比值进行加权。

2.3 系统发生树

我们用 PAUP4.0 软件的 P distance 方法计算遗传距离(表 1), 邻接法对 U83 基因构建果蝇科系统发生树(图 2)。所重建的树拓扑结构基本能反映果蝇科系统发生的大致过程: 同为果蝇亚属的 *D.mojavensis*, *D.virilis* 及 *D.grimshawi* 关系较近; 其他果蝇为水果果蝇亚属, 聚在一起。水果果蝇亚属中, 两个暗果蝇种组(Obscura species group)的姐妹种 *D.pseudoobscura*

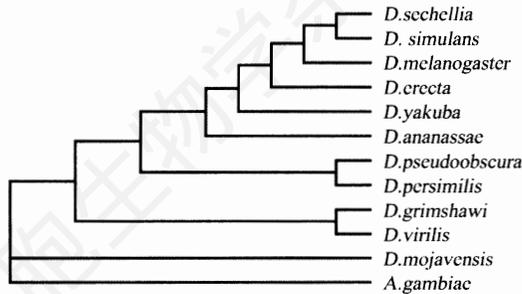


图 2 U83 基因序列系统进化分析

该系统进化树运用 PAUP4.0 软件中的邻接法构建。

和 *D.persimilis* 聚在一起; 水果果蝇亚属, 黑腹果蝇种的 6 种果蝇聚在一起, 并将黑腹果蝇亚种中的 5 种果蝇与 *D.ananassae* 区分开。另外, 姐妹种 *D.melanogaster*、*D.sechellia* 和 *D.simulans* 构成一簇, 与黑腹果蝇亚种中的 *D.yakuba* 和 *D.erecta* 区分开。新构建的系统发生树与传统方法所构建的果蝇科系统发生树^[9]虽然大致相同, 但还存在一定差别。例如, 传统方法构建的系统发生树认为 *D.erecta* 与 *D.yakuba* 是两个姐妹种, 它们在进化树中的地位相同, 而图 1 则显示 *D.erecta* 比 *D.yakuba* 进化程度更高。再如, 前者认为果蝇亚属与水果果蝇亚属是平行的两个分支, 且 *D.virilis* 与 *D.mojavensis* 为姐妹种, *D.grimshawi* 比这两个物种略早分化; 后者却显示 *D.grimshawi* 与 *D.virilis* 关系密切, 而 *D.mojavensis* 比包括水果果蝇亚属在内的其他果蝇更早分化。

为获得更多的序列信息, 我们将用来构建系统发生树的序列扩展至整个 U83 基因所在的内含子序列。通过 P distance 方法计算遗传距离(表 2), 邻接法构建系统发生树(图 3), 所得的系统发生树与传统

表 1 U83 基因序列差异百分比(p 值)

	1	2	3	4	5	6	7	8	9	10	11	12
1 <i>D.grimshaw</i>												
2 <i>D.virilis</i>	0.160											
3 <i>D.mojavensis</i>	0.214	0.188										
4 <i>D.melanoga</i>	0.201	0.164	0.211									
5 <i>D.sechelli</i>	0.201	0.164	0.211	0.000								
6 <i>D.simulans</i>	0.201	0.164	0.211	0.000	0.000							
7 <i>D.erecta</i>	0.187	0.151	0.210	0.012	0.012	0.012						
8 <i>D.yakuba</i>	0.232	0.182	0.191	0.073	0.073	0.073	0.085					
9 <i>D.ananassa</i>	0.186	0.162	0.186	0.083	0.083	0.083	0.071	0.133				
10 <i>D.pseudoob</i>	0.161	0.126	0.215	0.096	0.096	0.096	0.084	0.135	0.084			
11 <i>D.persimil</i>	0.161	0.126	0.215	0.096	0.096	0.096	0.084	0.135	0.084	0.000		
12 <i>A.gambiae</i>	0.507	0.508	0.470	0.492	0.492	0.492	0.480	0.465	0.469	0.487	0.487	

表 2 U83 所在内含子序列差异百分比(p 值)

	1	2	3	4	5	6	7	8	9	10	11	12
1 <i>D.ananassae</i>												
2 <i>D.erecta</i>	0.276											
3 <i>D.grimshawi</i>	0.436	0.423										
4 <i>D.melanogaster</i>	0.284	0.111	0.454									
5 <i>D.mojavensis</i>	0.423	0.393	0.296	0.417								
6 <i>D.persimilis</i>	0.301	0.290	0.397	0.274	0.419							
7 <i>D.pseudoobscura</i>	0.299	0.297	0.395	0.280	0.423	0.004						
8 <i>D.sechellia</i>	0.285	0.118	0.463	0.030	0.425	0.285	0.292					
9 <i>D.simulans</i>	0.283	0.126	0.450	0.030	0.417	0.278	0.285	0.023				
10 <i>D.virilis</i>	0.419	0.388	0.273	0.422	0.261	0.392	0.396	0.427	0.418			
11 <i>D.yakuba</i>	0.296	0.086	0.428	0.140	0.404	0.301	0.300	0.148	0.152	0.419		
12 <i>A.gambiae</i>	0.636	0.628	0.627	0.611	0.616	0.619	0.616	0.620	0.617	0.604	0.634	

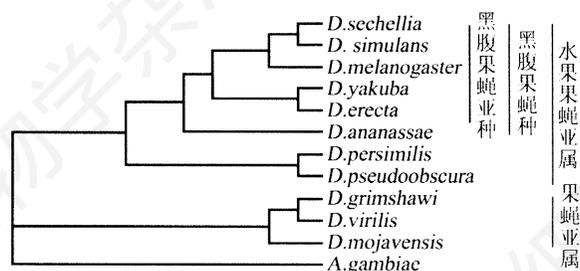


图3 U83 所在内含子序列系统进化分析

该系统进化树运用 PAUP4.0 软件的邻接法构建。

方法构建的果蝇科系统发生树拓扑结构几乎完全一致,只是在果蝇亚属3个物种间的关系上有所分歧:传统方法认为 *D. virilis* 与 *D. mojavensis* 为姐妹种, *D. grimshawi* 比这两个物种略早分化;本研究结果却显示 *D. grimshawi* 与 *D. virilis* 关系更密切。

3 讨论

snoRNA 广泛分布于真核生物和古细菌中,是一类重要的小分子 RNA。它们作用于 rRNA 的剪切、加工和成熟,指导 rRNA、snRNA、tRNA 碱基修饰,甚至与 mRNA 的编辑、端粒酶的稳定性的维持有关。正因为这类小分子 RNA 不容小觑的功能,近年来众多生物学家对它们表现出浓厚的兴趣。通过实验方法和计算机方法搜索,越来越多的 snoRNA 数据在 GenBank 上登录,多个与 snoRNA 相关的独立数据库相继建立,如 snoRNA 数据库(<http://lowelab.ucsc.edu/snoRNAdb/>),植物 snoRNA 数据库(http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home),小 RNA 数据库(<http://condor.bcm.tmc.edu/smallRNA/smallrna.html>),NONCODE(<http://noncode.bioinfo.org.cn/>)等。这些专设的二级数据库为我们提供了便捷的研究平台,突显了 snoRNA 数据量多,容易获得的优势。

U83 是众多 box C/D snoRNA 中的一员,本研究利用该基因构建果蝇科内各物种的种系发生树,并获得良好结果。利用 box C/D snoRNA 构建系统进化树是我们在 snoRNA 研究领域的新尝试,但这一尝试也具有一定理论基础和实践依据。通常用来构建系统发育树的序列有 16S/18S rDNA、rRNA ITS、mtDNA 12S、cyt b 等,同时具有易变区和保守区使这些分子成为构建分子系统发生树的理想序列。snoRNA 也和以上序列一样,同时具有易变区和保守区,这为用 snoRNA 序列构建分子系统进化树打下了理论基础。此外, Cervelli 等^[10]利用龟 U17 box H/

ACA snoRNA 构建了 16 种龟的系统进化树,结果与形态学结果一致,验证了用 box H/ACA snoRNA 构建分子系统发生树的可行性,这为利用 box C/D snoRNA 构建系统发生树提供了实践依据。

在割裂基因发现初期,人们认为内含子是基因组序列中的“垃圾”成分^[11]。但随着人类基因组序列和其他多种模式生物基因组序列的测序完成以及对非编码 RNA 研究的深入,人们发现内含子序列之中蕴藏着大量的 snoRNA 和 microRNA (miRNA)。它们在转录调控、RNA 的加工和修饰、细胞的分化发育等过程中发挥作用,成为当前生物学研究领域的热点和前沿。内含子序列也因而摆脱了“多余序列”的头衔,受到人们的广泛关注^[12]。编码 snoRNA 基因的内含子除 snoRNA 基因保守性强外,其余的序列变化速度较快,不受生存选择压力的影响。本文把比对序列延展至整个 U83 所在的内含子,克服了 U83 基因序列短、信息含量少的缺陷,使保守序列与可变序列有效结合,为系统进化树的构建提供了更多信息,建树结果与传统方法的建树结果几乎完全一致。

目前,国际上用于系统进化树构建的软件主要有 PAUP、PHYLP 和 MEGA 等,但一直以来,人们所研究的序列都以 rRNA、mtDNA 以及其他一些蛋白质序列为主,这些序列都比 snoRNA 序列长,使人们质疑用 PAUP 等软件构建短序列系统进化树的正确性。近来,随着人们对 miRNA 的深入研究,使 PAUP、PHYLP 等软件在构建仅 80 nt 的 miRNA 前体分子中得到肯定^[13,14]。这说明 PAUP 用来构建比 miRNA 前体序列长的 snoRNA 和 snoRNA 所在的内含子序列是适合的。此外,我们还用了适用于小分子 RNA 系统发生树构建的 Splits Tree4 软件^[15,16]分别对 U83 序列和 U83 所在的内含子序列构建系统发生树,得到与用 PAUP4.0 建树一致的结果(图略)。

本研究是用 box C/D snoRNA 基因序列构建系统发生树的首次尝试,其结果肯定了 U83 box C/D snoRNA 构建果蝇科种系发生树的价值,并为利用众多位于内含子中,序列较短的 snoRNA 基因构建系统发生树开创了方法。但 U83 仅是众多 box C/D snoRNA 中的成员之一,尽管 box C/D snoRNA 用于构建系统发生树有一定的理论基础,其普遍适用性还需要更多实验证据来支持。

参考文献 (References)

- [1] Storz G. *Science*, 2002, **296**: 1260

- [2] Tanaka-Fujita R *et al.* *RNA*, 2007, **13**: 811
[3] Makarova IuA *et al.* *Genetika*, 2007, **43**: 149
[4] Matera AG *et al.* *Nat Rev Mol Cell Biol*, 2007, **8**: 209
[5] Duga S *et al.* *Biochim Biophys Acta*, 2000, **1490**: 225
[6] Kumar S *et al.* *Brief Bioinform*, 2004, **5**: 150
[7] Thompson JD *et al.* *Nucleic Acids Res*, 1997, **25**: 4876
[8] Knight A *et al.* *Syst Biol*, 1993, **42**: 18
[9] Powell JR. *Progress and Prospects in Evolutionary Biology: The Drosophila Model*, Oxford: Oxford University Press, 1997
[10] Cervelli M *et al.* *J Mol Evol*, 2003, **57**: 73
[11] 张尚宏等. *中山大学学报(自然科学版)*, 1999, **38**: 49
[12] Luo Y *et al.* *Nucleic Acids Res*, 2007, **35**: 559
[13] Guddeti S *et al.* *Cell Res*, 2005, **15**: 631
[14] Tanzer A *et al.* *J Mol Biol*, 2004, **339**: 327
[15] Huson DH *et al.* *Mol Biol Evol*, 2006, **23**: 254
[16] Hertel J *et al.* *BMC Genomics*, 2006, **7**: 25

U83 box C/D snoRNA as a Molecular Marker in the Phylogeny of Drosophilidae

Ling Xu^{1,2}, Yu-Ping Luo¹, Dong-Gen Zhou¹, Si-Guang Li^{1*}

(¹College of Life Science, Nanchang University, Nanchang 330031, China; ²College of Life Science, Jiangxi Science and Technology Normal University, Nanchang 330013, China)

Abstract Based on FlyBase and U83 in *Drosophila melanogaster*, ten new U83 gene homologues were identified in Drosophilidae which were located in intron of rpl3 respectively. The homologous sequences were compared, and the molecular trees were reconstructed by neighbor-joining method of PAUP4.0 program using *Anopheles gambiae* as outgroup. The result reflected the evolutionary relationship in Drosophilidae roughly. Sequences were enlarged to the entire U83 located introns to get more information, and the phylogenetic tree constructed by those sequences was more correct compared with the former one. Genes of box C/D snoRNA were used for constructing phylogenetic tree for the first time, our study confirmed the good effect of U83 box C/D snoRNA as a molecular marker in the phylogenetic analysis of Drosophilidae.

Key words U83 snoRNA; intron; phylogenetic tree; Drosophilidae

Received: April 4, 2007 Accepted: June 15, 2007

This work was supported by the National Natural Science Foundation of China (No.30660042)

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