

综述与专论

短散在元件 (SINE) 的研究进展*

韩德民 周开亚¹⁾ 王义权

(南京师范大学遗传资源研究所, 南京 210097)

摘要 短散在元件 (SINE) 广布于真核生物, 是基因组中的可转移成分, 长约 100~ 500 bp, 拷贝数可达数百至数十万以上, 根据序列变异和鉴别位点常可分为若干家族和亚家族, 对基因组的复杂化、基因的钝化、新基因的产生、尤其是基因表达的调控都具有重要意义. 除了灵长类 Alu 和小鼠 B1 家族等少数来源于 7SL RNA, 大多为 tRNA 的衍生物, 由 tRNA 同源区、tRNA 无关区和富含 AT 区等组成. 在 tRNA 同源区含有 RNA 聚合酶 III 内部启动子成分. 其起源和进化有“主源基因模型”、“多源基因模型”、“寄生假说 (转座子模型)”和“平行转移假说”等. 以鲑科鱼类、鲸类及偶蹄类、高等灵长类和蛙类为例, 介绍了近年来短散在元件在分子系统发生研究以及在遗传作图和抗癌治疗上的应用.

关键词 短散在元件, 转移 RNA, 反转录子, 反转录转座, 进化, 系统发生

学科分类号 Q527

真核生物基因组的编码基因序列据估计仅约占 10%~ 20%, 其余主要为各种重复序列. 重复序列包括串联重复序列和散在重复序列两大类. 真核生物基因组中的散在重复序列主要有三种, 即短散在元件 (short interspersed element, SINE)、长散在元件 (long interspersed element, LINE) 和加工的逆转录假基因 (processed retropseudogenes). SINE 约占高等真核生物基因组的 1/3, 除少数如灵长类 Alu 和小鼠的 B1 来源于 7S RNA, 其余全部来源于 tRNA. 本文重点介绍后一类.

SINE 的结构、功能、起源和进化理论的研究, 是分子生物学和分子遗传学理论的重要组成部分, 对完善基因的概念、功能以及解决基因组“C 值矛盾”都有重要意义. 这方面的研究已经取得不少成果, 并提出了一些假说和模型, 有些研究还对 SINE 在生物系统发生上的意义作了探讨. 随着研究成果的积累, 尤其是近年来测序等技术的进步和序列资料在计算机网络上传递交流的快捷便利, 对 SINE 的研究获得了前所未有的快速发展.

1 SINE 的一般特征

SINE 的序列长度一般仅为 100~ 500 bp, 在基因组中的拷贝数可达数百至 10^5 以上, 散布或丛生在基因组中, 几乎可以分布于基因组中任何部位.

tRNA 起源的 SINE 具有特定的复合结构, 其序列从 5' 到 3' 端依次由 3 个区段组成: tRNA 相关区、tRNA 无关区和富含 AT 区 (包括多聚 A 尾巴). 和所有的转座子一样, 在序列插入位点两端也有 7~ 21 bp 的靶点倍生正向重复. 由于突变的积累, 该正向重复序列可变得不明显. tRNA 相关区序列与 tRNA 序列有高度的一致性 (约为 60%~ 80%), 内含两个不连续的保守区段, 为 RNA 聚合酶 III 内部启动子的主要成分, 末端三个碱基通常是 CCA. 因此, 该区段可能是来源于成熟的 tRNA. tRNA 无关区有认为源自某种 RNA 基因, 也有认为直接来源于相关的 LINE. 某些远缘物种的 SINE 如小鼠的 type 2、鲑鱼的 Sma、龟的 Pol III 和乌贼的 SK, 这一区域具有相似性, 特别有两段 GATCTG 和 TGG 高度保守, 两段间相隔 10~ 11 个核苷酸. 广泛分布于人及哺乳动物的 MIR (mammalian-wide interspersed repeat) 家族成员这一区域的中部也存在一段保守区, 称之为核心区 (core region). 这些保守区可能有识别和结合蛋白质的功能. Gilbert 等^[1] 认为该核心序列是一种古老的 SINE, 由它衍生了包括 MIR 在内的其他

* 国家自然科学基金资助项目 (39970094).

¹⁾ 通讯联系人.

Tel: (025) 3598328, E-mail: kyzhounj@jlonline.com

收稿日期: 2000-03-23, 修回日期: 2000-06-05

SINE 家族, 不只分布于哺乳类, 也广布于爬行类、鸟类、鱼类、以及软体动物等不同类群中, 其发生和扩散的历史可能超过 5.5 亿年. 根据 SINE 的结构特点, 一般认为它是由 RNA 介导的反转录转座 (retroposition) 形成的, 所以又称之为反转录子 (retroposon).

2 SINE 的多态性及其在生物界的分布

一个物种基因组中常分布有一种以上的 SINE. 同一种 SINE 的重复拷贝常有若干位点的差异, 有的有删节 (truncation) 现象, 因此常统称为 SINE 家族. 根据这些差异的大小和特异性的鉴别位点, 很多家族可划分为若干亚家族. SINE 虽然广布于真核生物, 但以哺乳动物为多. 就某一 SINE 来说, 有的分布非常广泛, 有的则仅分布于某一属中的少数几个种.

3 SINE 的反转录转座机理

3.1 SINE 与反转录病毒

SINE 的反转录转座过程很类似反转录病毒 (retrovirus) 的逆转录整合过程. 反转录病毒 U5 区序列中, 和 SINE 一样也多含有 GATCTG 和 TGG 这两个保守区段, 两区段间的间隔在两者中也基本一致, 而且大多数反转录病毒都以 tRNA^{Lys} 为引物整合到寄主基因组中, 而与 SINE 同源最多的 tRNA 也是 tRNA^{Lys}, 因此 tRNA 来源的 SINE 很可能也是以 tRNA 作为反转录引物. Kim 等^[2] 认为人类 SINE-R 就来源于人类 HERV-K (human endogenous retrovirus) 反转录病毒.

3.2 SINE 与 LINE

LINE 长 3~7 kb, 为 RNA 聚合酶 II 的转录物, 结构变异较大, 通常有 2 个开放阅读框, 编码逆转录酶及核酸内切酶, 末端有寡聚 T. 通过对多种 SINE 和 LINE 的比较研究, 发现 SINE 总是与特定的 LINE 相伴, 并且当某 LINE 失活时, 其相关的 SINE 也丧失反转录转座活力. SINE tRNA 无关区的 3' 端和分布于同物种的 LINE 的 3' 端有高度的同源性, 可能可以为同一种逆转录酶识别, 因此 SINE 的逆转录酶系可能是由 LINE 提供的. SINE 这种对 LINE 编码酶系的依赖有点象寄生的情形.

3.3 SINE 的整合

Jurka^[3] 通过对反转录转座插入位点附近 DNA 序列的比较分析, 认为 SINE 不是随机整合到基因

组中. 他发现 SINE 插入的交错切口附近通常有一个 5'-TTAAAA 六核苷酸序列, 可能是有关参与整合作用酶的识别位点, 并进一步推测该酶就是具有 AP 内切酶活性的逆转录酶.

4 SINE 的起源和进化

4.1 多拷贝的起源

如此多的 SINE 拷贝是如何发生的, 迄今尚无定论. 主要有两种假说:

4.1.1 “主源基因模型” (master source gene model): 认为多拷贝 SINE 的产生归因于单个 (或少数几个) 主源基因, 由主源基因序列通过反转录转座直接产生出所有的 SINE 拷贝, 而这些子代拷贝则不再有反转录转座活性. 由于“主源基因”作为不断产生新拷贝的“工作母机”而长期存在于基因组中, 所以该模型预测 SINE 在进化过程中对物种的生存延续有特定的功能.

4.1.2 “多源基因模型” (multiple source gene model): 与“主源基因模型”相反, 认为多个基因的同时扩增导致 SINE 的扩散. “多源基因模型”认为子代拷贝具有潜在的反转录转座活性, 因此造成庞大而没有多少功能的“自私的”DNA.

4.2 SINE 的扩散

SINE 伴随着物种基因组的进化而扩增, 因此似乎应该是垂直转移扩散的, 例如 MIR 广泛分布于哺乳动物, 以此解释就较为合理. 但有些远缘种分布有相似的 SINE, 则不好以此解释, 可能还有其他方式.

4.2.1 “寄生假说” (parasitism hypothesis), 又称“转座子模型” (transposon model): 原始成分的反转录转座受不同水平多种因素——如转座位点附近的染色体情况、甲基化作用、顺势作用启动子成分、转录作用因子、RNA 加工和多聚 (A) 代谢等的影响, 这些因素均影响反转录转座的结果, 因而在不同物种或生物类群中由于影响因素的不同, 各自的反转录转座频率亦不同. 在这一模型中, SINE 的作用象寄生.

4.2.2 “平行转移模型” (horizontal transfer model): 远缘物种有相似的 SINE, 可用两种平行转移的机制解释: 其一, 某种反转录病毒类似物在很短的进化时期内平行侵入这些远缘宿主基因组, 然后在各自的宿主体内独立形成 SINE. 已有多种证据支持这一假说. 根据反转录病毒现在的分布情况, 真正的反转录病毒应出现在脊椎动物甚至哺乳

动物出现之后, 因此 SINE 大量存在于哺乳动物基因组中是符合这一解释的. 其二, 原始 SINE 从一个宿主通过反转录子自身或 SINE DNA 转侵另一宿主物种. 水稻 pSINE1 的 tRNA 同源区像 tRNA^{Lys}, 但更像兔 SINE C 家族 tRNA 同源区, 似乎只能由此解释. Shimamura 等^[4]对鲸类和偶蹄类基因组中 SINE 的研究结论也支持这一假说. 最近还在植物中发现了果蝇的 Copia 家族. 烟草的 SINE TS 家族, 也可能由动物 SINE 超家族的一个成员水平转移入植物生成. 在这些转移过程中, 病毒和昆虫等可能起着媒介作用.

5 SINE 的功能

由于 SINE 的反转录转座使插入位点加入了新的 DNA 序列, 也改变了位点前后 DNA 序列之间的关系, 从而导致假基因和新基因的产生以及基因调控的变化, 这些都对基因组的功能产生很大的影响. 此外, SINE 在基因组和物种间的大量扩增和扩散, 加强了基因组基因间和物种间的信息交流, 从而增加了基因组的多态性, 并为新基因的产生和基因组的复杂化进化提供了广阔的基础, 其作用及重要性正逐渐被认识.

5.1 获得表达能力

并非所有 SINE 都是沉默的非编码序列, 不少 SINE 在反转录转座和进化过程中获得了编码蛋白质的能力. 例如偶蹄类的 CHR-1 家族中有的拷贝插入牛前列腺素受体 EP3 亚型基因的编码区而成为该编码基因的一部分^[5].

5.2 参与基因表达的调控

SINE 的作用可能更多地表现为对插入位点及其附近调控区的影响. 某些 SINE 含有特定的结构域, 供蛋白质识别和结合, 从而在转录和转录后加工的调控中发挥作用. 近来已有令人信服的证据表明, 小鼠 ID 中的第 9~72 核苷酸重复序列成分就起着控制基因表达的作用. Kang 等^[6]通过鼠的转基因试验, 发现带有 SINE 的外源基因的整合率较对照组可增加约 3 倍, 因此 SINE 极可能有增强同源重组的作用. 龟的 Pol III 家族中有一段一致序列 (GTGGTTTG) 与病毒增强子的核心序列完全相同, 因而可能是一种增强子成分.

热休克后, 鼠的 SINE B1 和 B2 的转录表达量有时在所有组织中都瞬时增加 40 倍, 类似于热休克蛋白. 乙醇的急性剂量也能瞬时增加 B1 和 B2 在肝脏的含量, 表明生理压力可增加 SINE 的表

达. B2 在所有组织中均表达, 属组成型表达, 而 B1 的表达则具有组织特异性, 表明这些 RNA 转录物具有正常的生理功能. 根据它的可诱导性, 所增加的表达是对生理压力的应答, 类似细胞的生理压力调控基因^[7].

6 在动物系统发生研究中的应用

SINE 的起源和进化无疑与生物进化历程密切相关. 和转座相比, SINE 的反转录转座更具有不可逆性而可作为生物进化过程中的标记. 越来越多的证据表明, 许多 SINE 在生物类群中的分布具有特异性. 因此, 近年来已有一些研究者利用 SINE 来探讨有关动植物类群的系统发生, 以推动分子系统学的发展.

6.1 对鲑科鱼类和鲸类的研究

鲑科 (Salmonidae) 鱼类种类繁多, 大多分化于近期, 是研究系统发生的好材料. Okada 研究组首先从鲑科鱼类基因组中分离并分析了 3 个 SINE 家族 Sma I、Fok I 和 Hpa I, 发现了 SINE 在鲑科鱼类属种间的特异性分布. 此后, 以 12 种鲑科鱼类为材料, 分析了 Hpa I 家族中的几个亚家族, 结果推论这 12 种鲑鱼分为 3 支并重建了系统发生树. 通过 SINE 拷贝数和特异性核苷酸鉴别位点的分析, 该研究组还探讨了 SINE 在鲑科鱼类进化史中的水平扩增历程, 并据此重建了太平洋和大西洋鲑科鱼类的系统发生树^[8,9].

该研究组还以 SINE 作为鲸类起源于偶蹄类的证据. 此前古生物学、形态学和生物分子研究均认为鲸目 (Cetacea) 与偶蹄目 (Artiodactyla) 亲缘关系最近, 但对偶蹄目是否是单系仍有疑问. 对鲸目和偶蹄目 2 个 SINE 家族 (CHR-1, CHR-2) 9 个成员的研究表明, 鲸目、反刍亚目 (Ruminantia) 和河马 (*Hippopotamus amphibius*) 构成单系, 这一单系不包括偶蹄目中的骆驼和除河马外的其他猪亚目 (Suiformes), 因此偶蹄目应是并系的^[4,10].

6.2 对高等灵长类的研究

Kim 等^[11]发现 SINE-R 是来源于人类 HERV-K 反转录病毒的反转录子家族, 他们用来源于人脑的 SINE-R 序列为引物作 PCR 扩增, 从黑猩猩得到了 27 个序列, 从大猩猩得到了 16 个序列, 与已知的人以及猩猩和长臂猿的序列进行比较, 发现它们之间有广泛的重叠, 揭示了人类进化过程中的多起源性, 认为 SINE-R 家族在灵长类进化过程中可

能仍在继续活跃转座。分析还表明, 人第 7 和第 17 染色体上的 SINE-R 成分与类人猿互相关联, 这些反转录子的各个亚家族在人类进化过程中独立进化^[2]。

6.3 对蛙类的研究

Bucci 等^[12]研究了 Rana/Pol III 家族在 Rana 属 6 种及 2 个杂交种的分布, 研究结果与已有的遗传进化分析一致, 提出杂交遗传机制是使 Rana/Pol III 从欧洲中部和北部的蛙类向伊比利亚的 R. perezii 水平转移的因素之一。

7 在其他方面的应用

SINE 可以用来建立 PCR 介导的多位点指纹图谱。Kostia 等^[13]用根据牛等有蹄类 Bov-tA (bovine tRNA pseudogene coupled to an A element) 设计的引物扩增并分析了 6 种偶蹄类 SINE 的指纹图谱。根据指纹数据的特征和距离重建的分子系统发生树, 与其他方法重建的一致。

顺氯氨铂 (cis-dichlorodiammine platinum II) 是最有效的抗癌剂之一, 但其对健康组织尤其是肾的毒性限制了它的使用, 如能在治疗过程中监控顺氯氨铂-DNA 加合物, 则可以解决增加化疗效果并减少副作用的问题。Wang 等^[14]发现了一种高灵敏度分析方法, 依赖于哺乳动物基因组中高拷贝保守性的散在元件, 可在体内外检测顺氯氨铂加合物。他们用鼠的血细胞试验, 发现鼠的 B1 可成功地应用于这一方法。

参 考 文 献

- Gilbert N, Labuda D. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci USA, 1999, **96** (6): 2869~ 2874
- Kim H S, Crow T J. Phylogenetic relationships of a class of hominoid-specific retro elements (SINE-R) on human chromosomes 7 and 17. Ann Hum Biol, 2000, **27** (1): 83~ 93
- Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA, 1997, **94** (5): 1872~ 1887
- Shimamura M, Abe H, Nikaido M, et al. Genealogy of families of SINEs in cetaceans and artiodactyls: the presence of a huge superfamily of tRNA (Glu)-derived families of SINEs. Mol Biol Evol, 1999, **16** (8): 1046~ 1060
- Shimamura M, Nikaido M, Ohshima K, et al. A SINE that acquired a role in signal transduction during evolution. Mol Biol Evol, 1998, **15** (7): 923~ 925
- Kang Y K, Park J S, Lee C S, et al. Efficient integration of short interspersed element-flanked foreign DNA via homologous recombination. J Biol Chem, 1999, **274** (51): 36585~ 36591

- Li T, Spearow J, Rubin C M, et al. Physiological stresses increase mouse short interspersed element (SINE) RNA expression *in vivo*. Gene, 1999, **239** (2): 367~ 372
- Murata S, Takasaki N, Saitoh M, et al. Details of retropositional genome dynamics that provide a rationale for a genetic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). Genetics, 1996, **142** (3): 915~ 926
- Hamada M, Kido Y, Himberg M, et al. A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the Sma I family of repeats: possible evidence for the horizontal transfer of SINEs. Genetics, 1997, **146** (1): 355~ 367
- Shimamura M, Yasue H, Ohshima K, et al. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature, 1997, **388** (14): 666~ 670
- Kim H S, Crow T J. Phylogenetic relationships of a class of hominoid-specific retro elements (SINE-R) on human chromosomes 7 and 17. Ann Hum Biol, 2000, **27** (1): 83~ 93
- Bucci S, Raghianti M, Mancino G, et al. Rana/Pol III: a family of SINE-like sequences in the genomes of western Palearctic water frogs. Genome, 1999, **42** (3): 504~ 511
- Kostia S, Ruohonen-Lehto M, Vainola R, et al. Phylogenetic information in inter-SINE and inter-SSR fingerprints of the Artiodactyla and evolution of the Bov-tA SINE. Heredity, 2000, **84** (1): 37~ 45
- Wang G, Hallberg L M, Englander E W. Rapid SINE-mediated detection of cisplatin: DNA adduct formation *in vitro* and *in vivo* in blood. Mutat Res, 1999, **434** (2): 67~ 74

Recent Achievements of Research on Short Interspersed Elements. HAN De-Min, ZHOU Kai-Ya, WANG Yi-Quan (Institute of Genetic Resources, Nanjing Normal University, Nanjing 210097, China).

Abstract SINEs (short interspersed elements) are mobile elements of approximately 100~ 500 bp that are often present as more than 10^5 copies per genome. They can be divided into families and subfamilies according to the mutational or diagnostic sequence loci. SINEs are widespread in eukaryotic genomes and create additional sequence combinations through dispersal and exchange of genetic information. So, they are believed to be of major importance in creating genetic diversity, gene inactivity, new gene, and especially in gene expression and gene regulation. Almost all SINEs reported to date are derived from tRNAs, with the exception of the primate Alu and the rodent B1 families, which are derived from 7SL RNA. The tRNA-derived SINEs have a composite

structure, with a region homologous to a tRNA, a middle tRNA-unrelated region, and a terminal AT-rich region. Each SINE contains an internal promoter for RNA polymerase III and lacks open reading frames. "Master source gene model", "multiple source gene model", "transposon model (parasitism model)", and "horizontal transfer model" have been postulated to explain the origin and evolution of

SINES. The observation of their molecular phylogeny was made with examples of salmon, whales and artiodactyls, hominoid primates, and western Palearctic water frogs. Their application to genetic profiles and antitumor technique was also introduced. **Key words** Short Interspersed Elements (SINE), tRNA, retroposon, retroposition, evolution, phylogeny

G 蛋白在细胞机械感受机制中的作用*

冯 兵 王德文 何作云¹⁾

(军事医学科学院放射医学研究所, 北京 100850)

摘要 机械刺激对生物器官、细胞的生理活动具有重要的影响。虽然生物细胞对各种机械作用具有各自不同的反应形式, 但大多具有类似的机械感受机制, 即首先由细胞膜、整合素-细胞骨架感受, 在激活 G 蛋白等信号分子后, 进一步引起生理、病理反应。因此, G 蛋白可能是完成机械能向化学能转化的关键环节。

关键词 机械受体, 信号传导, 心肌细胞, 血管内皮细胞

学科分类号 R339.17, Q689

大量研究发现包括器官运动 (如肢体运动、心搏、血液循环)、细胞间粘附以及细胞内的应力等不同性质、不同程度、不同持续时间的机械刺激在器官/细胞发育、分化以及生长适应过程中均有着重要的生理、病理作用^[1]。近年来, 对于多种类型细胞 (主要是心肌细胞、血管内皮细胞) 受到机械刺激后的细胞内信号分子及生理、病理效应有了比较清晰的认识, 但是由于技术水平及手段的限制, 对细胞初始机械感受机制了解不多^[1,2]。已有实验提示 G 蛋白是心肌细胞和内皮细胞等机械刺激初始反应中的重要环节^[2-4], 鉴于 G 蛋白在受体介导的信号传导机制中处于调控中枢地位, 因此, 阐明 G 蛋白在细胞机械刺激感受中的作用可能是寻找机械感受受体的一条途径。本文重点概述 G 蛋白在心肌细胞和血管内皮细胞机械感受机制中的作用。

1 细胞信号跨膜转导机制研究概况

细胞外信号主要包括细胞外的物理、化学信号以及细胞自身生物大分子的结构信号, 后者包括蛋白质的形状和构象、寡糖链与膜蛋白、膜脂相连而成的天线结构等。

对于细胞外化学信号跨膜转导的认识已比较清楚, 即: 细胞外信号-细胞膜受体-G 蛋白 (或非 G 蛋白)-细胞内第二信使, 受体是刺激-反应偶联信号跨膜通路中的关键部分。

至今, 对于细胞外物理信号, 尤其是机械信号的感受和跨膜传导机制了解甚少。Ingber 等^[5]提出了一个细胞机械刺激信号传递模型, 模型的基础是细胞骨架网络构成的刚性结构和力学作用的转导 (transduction), 而非传递 (transmission, 即化学信号的移位和级联反应), 力学转导的终极点必须能将机械能转化为化学能。与化学信号跨膜转导模型一样, 它们都强调 G 蛋白的作用。大 G 蛋白通过对其亚基上氨基酸残基的脂化修饰作用而锚定在细胞膜上, 从而为其接受细胞膜结构信号提供了结构基础。而小 G 蛋白也可通过细胞外基质-整合素-细胞骨架接受细胞外机械信号, 如 Rab、Arf、Sar 与细胞内蛋白质和囊泡转运有关, Ras 在介导 MAPK 磷酸化级联反应方面发挥关键作用, 而

* 国家自然科学基金资助项目 (39600041)。

¹⁾ 第三军医大学新桥医院心内科, 重庆 400037

Tel: (010) 66931236, E-mail: fengxb12@yahoo.com

收稿日期: 1999-08-27, 修回日期: 1999-12-24