中国科学: 生命科学 2025 年 第 55 卷 第 8 期: 1575 ~ 1590
SCIENTIA SINICA Vitae lifecn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
评 述 双清论坛 跨尺度RNA调控与成像专题

R2逆转座子系统中RNA的结构与功能研究进展

谭顺青[1,2,3,4*], 杨韵[1,2,3,4], 杨启羽[1,2,3,4], 刘俊杰[1,2,3,4*]

1. 清华大学生命科学学院, 北京 100084
2. 清华大学, 北京生物结构前沿研究中心, 北京 100084
3. 清华大学, 清华大学-北京大学生命科学联合中心, 北京 100084
4. 清华大学, 膜生物学国家重点实验室, 北京 100084
* 联系人, E-mail: tansq21@mails.tsinghua.edu.cn; junjiegogoliu@tsinghua.edu.cn

收稿日期: 2025-05-06; 接受日期: 2025-07-28; 网络版发表日期: 2025-08-06
中国农业农村部、科技创新2030——重大项目(批准号: 2022ZD0213900)、国家自然科学基金(批准号: 32425030)、新基石科学基金会和北京市自然科学基金(批准号: QY23135)资助

摘要 如何实现基因大片段的高效、精准插入, 是基因编辑领域的重要难题之一. 尽管基于CRISPR-Cas系统开发的各类基因编辑技术已实现碱基水平的插入、删除和替换, 但是大片段DNA整合的效率和精度仍然是亟待突破的瓶颈. 逆转座子, 一类具有大片段插入活性的天然遗传元件, 为解决这一难题提供了极具潜力的可行方案. 近年来, 逆转座子的机制研究和工具开发方面均取得了重大的突破. 本文聚焦R2逆转座子, 系统梳理了R2从关键功能元件鉴定、分子机理深入解析到基因编辑工具开发的发展历程; 重点剖析了R2系统中RNA组分的核心结构与功能, 并归纳了将R2改造为基因大片段插入工具的突破性策略和思路; 期望深化对逆转座子生物学功能的理解, 为逆转座子的工具化改造提供理论参考.

关键词 RNA结构, 基因编辑, 基因大片段插入, non-LTR逆转座子, R2逆转座子

在过去的二十年里, 基因编辑领域经历了革命性突破. 以CRISPR-Cas9为代表的核酸酶系统通过诱导DNA双链断裂(double-strand break, DSB), 依赖细胞内非同源末端连接(nonhomologous end-joining, NHEJ)或同源定向修复(homology directed repair, HDR)通路, 可以实现高效的基因敲除和敲入[1]. 基于此开发的单碱基编辑器(base editor, BE)[2]、引导编辑器(prime editor, PE)[3]以及可编程靶向添加技术(programmable addition via site-specific targeting elements, PASTE)[4]等, 通过不同机制实现了碱基替换、小片段插入或删除等精准编辑. 然而, 当目标编辑片段长度超过1 kb时, 上述技术面临编辑效率低、精准度下降等问题, 并受到工具分子尺寸大、大片段DNA模板递送困难的限制[5,6]. CRISPR偶联转座酶(CRISPR-associated transposases, CAST)技术显著提升了可编辑长度, 能够完成大于10 kb的基因插入. 但是, CAST技术同样受限于大片段DNA模板的递送问题[7]. 因此, 开发稳定、高效的基因大片段插入工具, 仍然是基因编辑领域亟待突破的技术

引用格式: 谭顺青, 杨韵, 杨启羽, 等. R2逆转座子系统中RNA的结构与功能研究进展. 中国科学: 生命科学, 2025, 55: 1575–1590
Tan S Q, Yang Y, Yang Q Y, et al. Progress of research on the structure and function of RNA in R2 retrotransposon (in Chinese). Sci Sin Vitae, 2025, 55: 1575–1590, doi: 10.1360/SSV-2025-0147

© 2025《中国科学》杂志社 www.scichina.com
https://www.sciengine.com/doi/10.1360/SSV-2025-0147

瓶颈.

在此背景下, 逆转座子这一天然生物元件的工程化改造展现出独特的优势. 逆转座子是生物体基因组中广泛存在的可移动遗传元件(mobile genetic element, MGE). 以RNA为中间体, 利用自身编码的逆转录酶(reverse transcriptase, RT), 逆转座子可将长达数kb甚至数十kb的遗传信息整合至基因组[8]. 因此, 与人工设计的PE和PASTE等CRISPR衍生系统和CAST技术相比[9], 逆转座子有望规避递送难题, 展现更加优越的性能.

自Barbara McClintock于20世纪40年代发现转座现象以来, 大量的MGE被陆续鉴定和报道[10,11]. 根据转座过程是否依赖RNA中间体, MGE可分为DNA转座子和逆转座子两大类[12]. 其中, 逆转座子可细分为长末端重复(long terminal repeat, LTR)逆转座子与非长末端重复(non-long terminal repeat, non-LTR)逆转座子(图1A)[8]. LTR逆转座子(如酵母中的Ty元件)通过病毒样颗粒(virus-like particle, VLP)完成转座, 其整合过程依赖整合酶(integrase, IN)对LTR序列的特异性识别, 改造和开发受到LTR序列的限制[13]. 相比之下, non-LTR逆转座子(如R2和L1元件)采用靶点引导的逆转录(target-primed reverse transcription, TPRT)机制进行转座, 即核酸内切酶(endonuclease, EN)结构域首先在基因组靶标位点产生单链切口, 随后RT结构域以暴露的3′-羟基(3′-OH)作为引物起始逆转录, 然后EN结构域切割基因组第二条链, 最后完成第二条链合成及缺口修复, 将自身遗传信息整合至宿主基因组中(图1B和C)[14~16]. 因此, non-LTR逆转座子独特的TPRT机制使得靶标位点的选择不受LTR的限制, 更具有应用潜力.

尽管大多数non-LTR逆转座子几乎随机地整合到宿主基因组中(如L1识别“TTTT/AA”保守序列), 但是仍有许多non-LTR逆转座子整合到特定序列中(如R2专一地识别28S rDNA基因内的高度保守区域), 天然的靶向精确性为人工改造提供了理想的框架[17,18]. 近年来, 随着冷冻电子显微镜(电镜)技术的发展, 多个不同转座状态的R2复合物的结构被解析, 揭示了R2蛋白与模板RNA和靶标DNA之间的相互作用细节, 阐明了R2蛋白特异性选择其模板RNA和靶标DNA的分子机制[19~23]. 这些基础研究的突破, 推动了工程化R2逆转座子工具的快速发展, PRINT(precise RNA-mediated insertion of transgenes)和STITCHR(site-specific target-primed insertion through targeted CRISPR homing of retroelements)等多项基于R2的基因大片段插入工具被先后提出[24~28]. 因此, 本文以R2逆转座子为例, 介绍R2逆转座子从发现, 到阐明, 再到应用的研究历程, 重点关注人们对R2系统中起关键作用的RNA的结构与功能研究, 并整理了工程化改造R2的进展和局限, 期望为未来更好地理解和利用逆转座子指明方向.

1 R2逆转座子的鉴定和分类

1.1 R2逆转座子的发现

在黑腹果蝇(*Drosophila melanogaster*, *Dm*)的rDNA基因簇中, 超过一半的rDNA存在非编码序列的插入, 其中, 28S rDNA存在两种不同长度的插入序列(分别为type1和type2)(图2A)[29]. 1981年, Dawid和Rebbert[30]对这两类插入序列与28S rDNA之间的边界进行了分析并发现, 与完整的28S rDNA序列相比, type1插入序列的两侧存在14 bp的rDNA序列重复, 而type2插入序列的两侧则存在2 bp的rDNA序列缺失. 1985年, Eickbush和Robins[31]在家蚕(*Bombyx mori*, *Bm*)的rDNA基因簇中同样观察到这两类插入序列. 两侧边界序列的不同, 暗示了这两类插入序列可能由不同的机制所介导. 因此, 1987年, Xiong和Eickbush[32]正式将这两类插入序列命名为R1(type1)和R2(type2). 此后的研究证明, R1属于脱嘌呤/嘧啶核酸内切酶(apurinic/apyrimidinic endonuclease, APE)类型non-LTR逆转座子, 而R2属于类限制性核酸内切酶(restriction enzyme-like endonuclease, RLE)类型non-LTR逆转座子[33].

1987年, Burke等人[11]对家蚕进行了基因组印迹和测序分析, 确定了R2基因的完整序列: R2*Bm*基因全长约4.2 kb, 包含5′端非翻译区(5′-untranslated regions, 5′-UTR)、一个开放阅读框(编码一个约1100个氨基酸组成的蛋白质, 其中RT结构域和L1具有同源性)和3′端非翻译区(3′-untranslated regions, 3′-UTR). R2*Bm*在家蚕基因组中大约存在25个拷贝, 至少20个拷贝插入在28S rDNA的同一个位点, 这意味着R2*Bm*是一类插入位点特异的non-LTR逆转座子. 1991年, Jakubczak等人[34]证明, 大多数昆虫的28S rDNA中均有R2的插入, 插入频率从28S rDNA的几个百分点到超过50%不等. 此后的研究表明, R2至少分布在九个动物门类中, 包含鸟

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

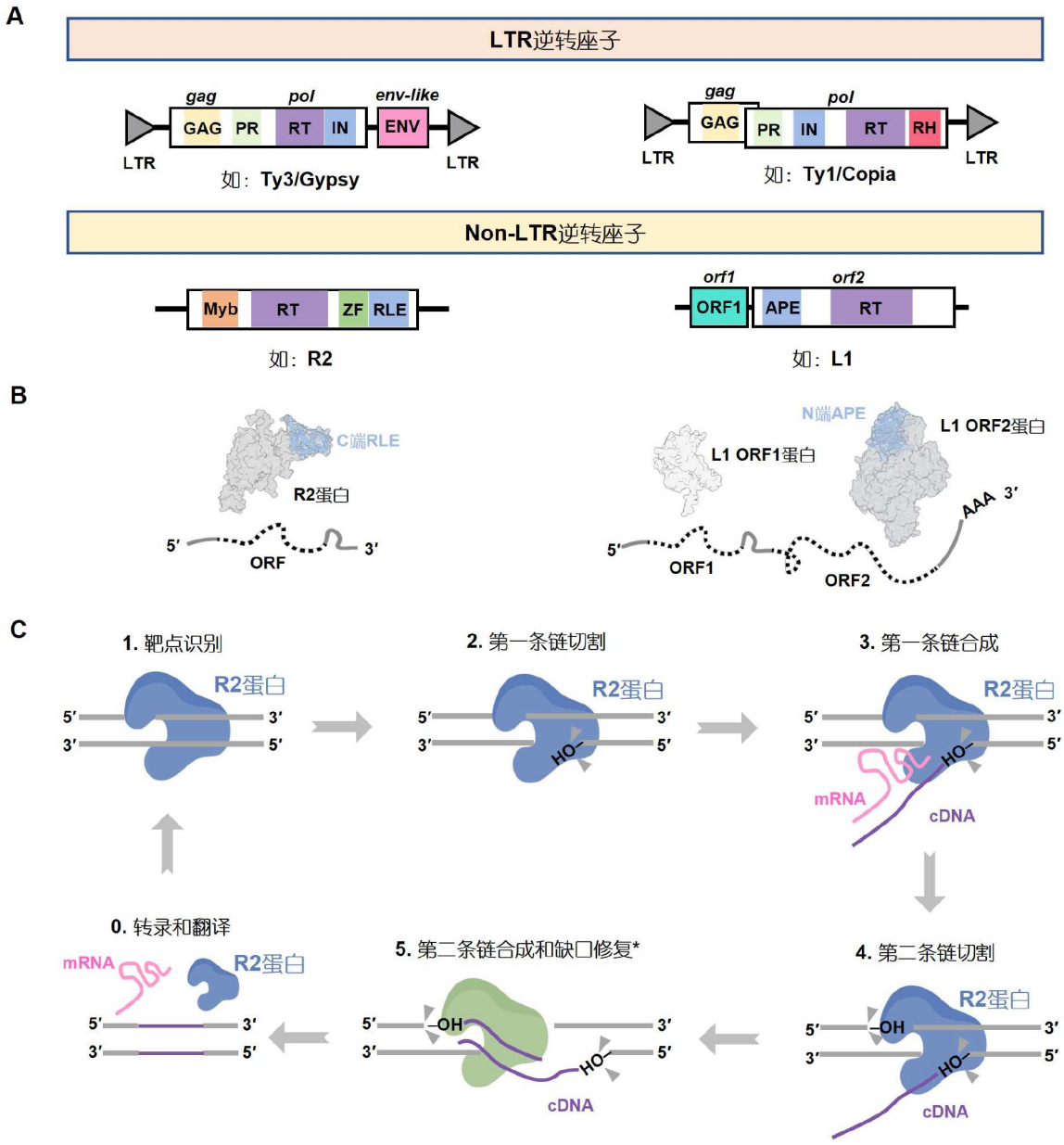


图 1　逆转座子的分类和TPRT模型. A: 代表性的LTR逆转座子和non-LTR逆转座子. *gag*: 群特异性抗原基因; *pol*: 聚合酶基因; *env-like*: 类包膜基因; *orf1*: 开放阅读框1基因; *orf2*: 开放阅读框2基因; GAG: *gag*基因编码蛋白; PR: *pol*基因编码蛋白酶; RT: 逆转录酶; IN: 整合酶; ENV: 包膜蛋白; RH: 核糖核酸酶H; Myb: c-myc样DNA结合结构域; ZF: 锌指结构域; RLE: 限制内切酶类核酸酶; ORF1: 开放阅读框1编码蛋白; APE: 脱嘌呤/嘧啶核酸内切酶; B: R2逆转座子与L1逆转座子; C: non-LTR逆转座子的“复制-粘贴”机制, *可能由细胞修复通路相关蛋白介导. mRNA: 信使RNA; cDNA: 互补DNA

Figure 1　Classification of retrotransposons and the TPRT model. A: Representative LTR retrotransposons and non-LTR retrotransposons; *gag*: group-specific antigen gene; *pol*: polymerase gene; *env-like*: envelope-like gene; *orf1*: open reading frame 1 gene; *orf2*: open reading frame 2 gene; GAG: *gag*-encoded protein; PR: *pol*-encoded protease; RT: reverse transcriptase; IN: integrase; ENV: envelope protein; RH: ribonuclease H; Myb: c-myc-like DNA-binding domain; ZF: zinc finger domain; RLE: restriction-like endonuclease; ORF1: *orf1*-encoded protein; APE: apurinic/apyrimidinic endonuclease; B: R2 retrotransposon and L1 retrotransposon; C: the “copy-and-paste” mechanism of non-LTR retrotransposons, * possibly accomplished by the cellular repair pathway. mRNA: message RNA; cDNA: complementary DNA

类、昆虫、鱼类和线虫类等[35]. 在同一物种基因组中, 除5′端截短外, 不同R2拷贝之间的突变积累很少[36]. 这些数据表明, R2是一个活跃的、物种分布广泛的位点特异型non-LTR逆转座子.

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

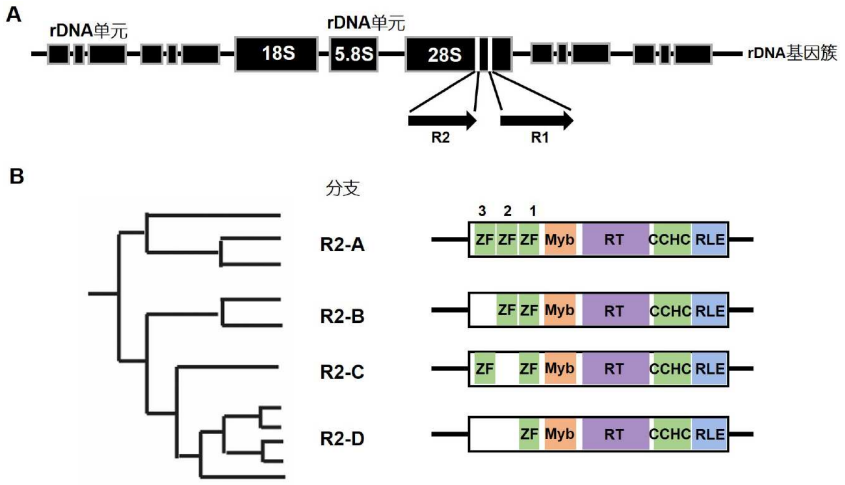


图 2 R2逆转座子的插入位点和分类. A: R2插入位点的示意图; B: R2的分类和结构域排布, 图片修改自文献[39](开放获取). CCHC: $CX_2CX_4HX_4C$锌指

Figure 2 Insertion sites and classification of R2 retrotransposons. A: Schematic diagram of the insertion site; B: classification and domain structures of R2, adapted from ref. [39] (Open Access)

1.2 R2逆转座子的分类

在不同物种来源的R2中, 尽管5′-UTR和3′-UTR的长度差异较大, 但是其编码蛋白具有高度的保守性, 均包含位于N端的ZF基序和Myb结构域、RT结构域和位于C端的RLE结构域[37]. 2005年, Kojima和Fujiwara[38]对R2蛋白的系统发育分析表明, R2在进化上可以细分为四个分支, 分别命名为R2-A, R2-B, R2-C和R2-D, 它们在超过8.5亿年前从一个共同祖先演化而来(图2B). 四个分支在N端ZF基序的数量上存在差异: R2-A分支包含3个ZF基序, 从N端起, 分别为$CX_2CX_{12}HX_4H$ (ZF3), $CX_2CX_{12}HX_3C$(ZF2), $CX_2CX_{12}HX_4H$(ZF1); R2-B分支包含2个ZF基序, 缺少ZF3; R2-C分支包含2个ZF基序, 缺少ZF2; R2-D分支包含1个ZF基序, 缺少ZF2和ZF3, 最早鉴定的R2*Dm*和R2*Bm*均属于R2-D分支[39,40]. ZF基序通常具有识别核酸的功能, R2不同分支在ZF基序数量和种类上的差异暗示了它们可能具有不同的靶向机制, 随演化过程逐渐被固定在了28S rDNA中[39].

2 R2逆转座子中的RNA元件

在non-LTR逆转座子的“复制-粘贴”扩增过程中, RNA中间体十分重要. RNA中间体由non-LTR逆转座子转录而来, 具有多重角色(图1B和C): 一方面, RNA中间体携带蛋白编码信息, 是蛋白翻译的模板(“复制”过程); 另一方面, RNA中间体携带non-LTR逆转座子的完整序列, 是TPRT的模板(“粘贴”过程)[15]. 在R2逆转座子中, RNA中间体上存在5′-UTR, 5′-RNA和3′-UTR三段具有保守结构和功能的元件, 发挥着不同的作用, 共同保证了R2的正常扩增[41].

2.1 5′-UTR的结构与功能研究

RNA转录本的产生(即“复制”过程)是逆转座子完整转座周期中非常重要的一个环节, 关系到后续的蛋白质翻译和逆转录. 大多数non-LTR逆转座子, 如L1逆转座子, 通过自身编码启动子来起始转录[42]; 然而, 部分non-LTR逆转座子自身不编码启动子, 而是借用插入位点上游的启动子进行转录[33,43]. 1999年, George和Eickbush[44]的实验证明了R2逆转座子不含启动子序列. 2010年, Eickbush等人发现了位于R2逆转座子5′端的非编码序列5′-UTR具有自切割活性, 并验证了生化测试中的自切割产物与拟果蝇(*Drosophila simulans*, *Ds*)中分离的R2成熟转录本具有一样的5′末端序列, 证明了R2利用5′-UTR的自切割活性, 从rRNA-R2转录本前体中加工出R2成熟转录本(图3A)[45]. 2013年, 他们进一步验证了5′-UTR结构和功能的保守性, 尽管5′-UTR在不同物种中的长度存在差异, 自切割位点也不尽相同, 但是它们均具有和HDV(hepatitis delta virus)

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

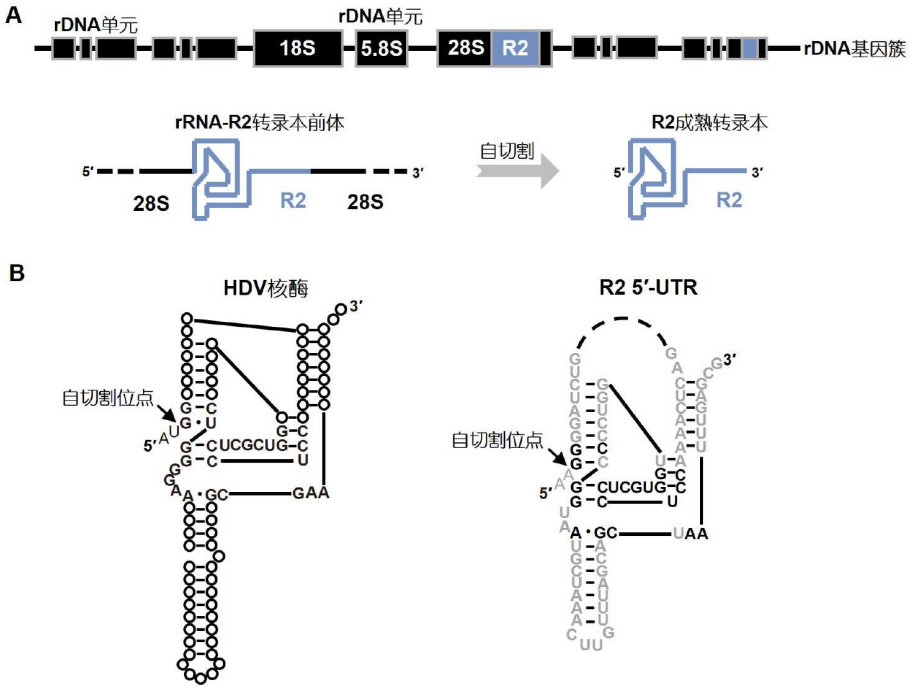


图 3　5′-UTR的结构与功能. A: R2转录本的产生和加工示意图; B: 5′-UTR与HDV核酶的二级结构比较, 图片修改自文献[45] (开放获取)

Figure 3　The structure and function of 5′-UTR. A: Schematic diagram of generation and processing of R2 transcript; B: conserved secondary structure of 5′-UTR and HDV ribozyme, adapted from ref. [45] (Open Access)

核酶相似的二级结构和自切割活性(图3B)[46].

2.2　3′-UTR的结构与功能研究

1993年, Luan等人[14]首次建立了分析家蚕来源的R2*Bm*酶学性质的实验体系, 检测了R2*Bm*蛋白的RLE和RT结构域的生化活性. 位于R2开放阅读框下游的3′-UTR序列是起始TPRT的必需元件, R2严格地从3′-UTR序列的最末端开始合成互补DNA(complementary DNA, cDNA), 并且3′-UTR末端与DNA底物之间无需任何同源互补配对序列, 延长或缩短3′-UTR的长度均会降低逆转录的活性(图4A)[15,47]. 2023年, Deng等人[20]解析的R2*Bm*冷冻电镜结构显示, RT结构域活性口袋中的第323位赖氨酸、第324位半胱氨酸、第406位丝氨酸和第446位精氨酸通过氢键相互作用, 精准识别并捕获了3′-UTR的末端. 然而, 3′-UTR中并未发现自切割核酶的保守序列, 目前尚不清楚细胞中与rDNA共表达的R2是否能从转录本中加工出成熟的3′末端. 凝胶迁移实验表明, 3′-UTR可以增强R2*Bm*蛋白与DNA底物之间的结合强度[48,49], 并增强第一条链切割活性[20,47]. 因此, 3′-UTR除了作为起始TPRT的模板, 还参与调控R2*Bm*的核酸内切酶活性, 发挥着双重功能.

1997年, Mathews等人[50]在测定TPRT活性时发现, R2*Bm*蛋白可以识别黑腹果蝇来源的3′-UTR并合成cDNA. 2004年, Ruschak等人[51]分析了不同物种来源的3′-UTR序列, 进一步证明了一级序列无显著相似性的3′-UTR之间在二级结构上具有保守性, 均存在大量的茎环(stem loop)结构. 以R2*Bm*为例, 基于寡核苷酸微阵列和SHAPE等技术的二级结构测定表明, R2*Bm*的3′-UTR由6个茎环构成(图4B)[52].

近三年内, 多个团队解析的冷冻电镜高分辨结构共同揭示了3′-UTR被R2蛋白结合并起始TPRT的分子细节[19~22]. R2*Bm*的结构显示, R2蛋白通过RNA结合结构域(即RB结构域, 同NTE-1)和RT活性口袋特异性地识别3′-UTR的核心结构(茎环2和茎环5), 其中, 茎环2的“G31-G39”链呈现明显的弯折(图4B)[19,20]. 在斑胸草雀(*Taeniopygia guttata*, *Tg*)和平胸龟(*Platysternon megacephalum*, *Pm*)来源的R2的冷冻电镜结构中, 茎

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

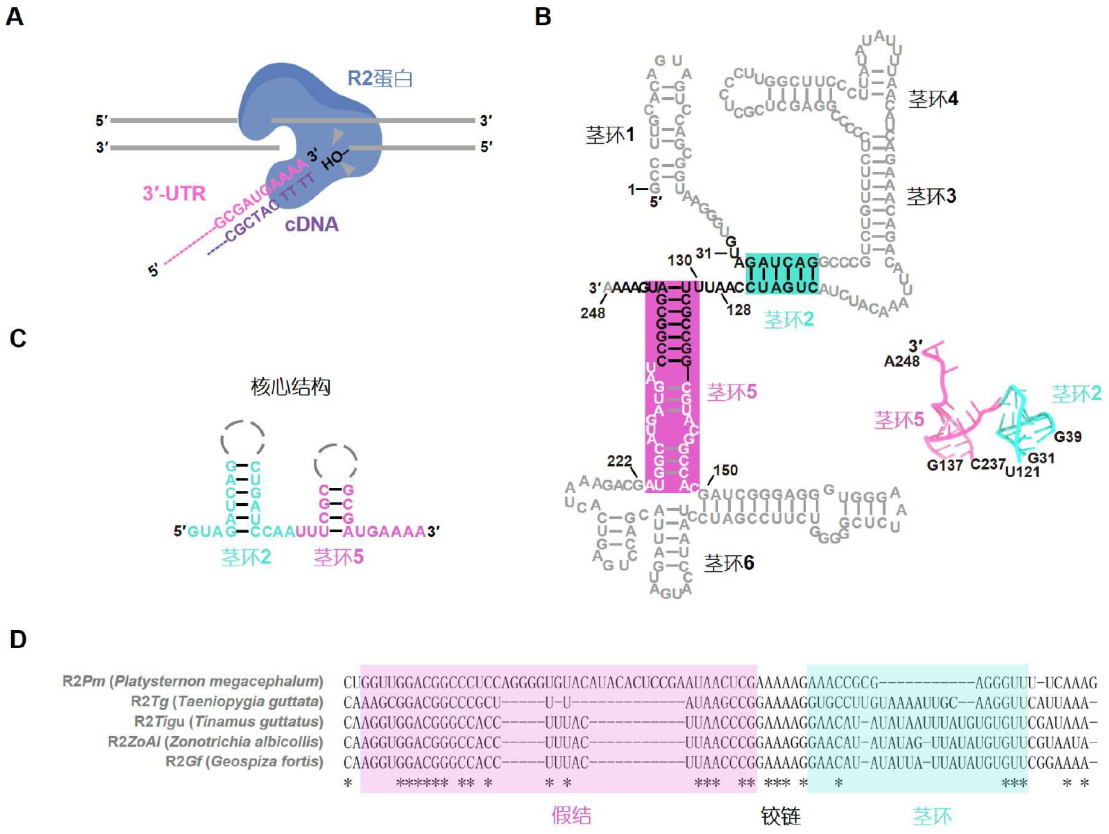


图 4 3′-UTR的结构与功能. A: 3′-UTR起始TPRT的示意图; B: 3′-UTR的二级结构保守性(PDB: 8IBW), 图片修改自文献[20](开放获取); C: 3′-UTR的核心区域, 图片修改自文献[19,21](开放获取); D: R2-A分支的3′-UTR中保守的“假结-铰链-茎环”结构, 图片修改自文献[21](开放获取)

Figure 4 The structure and function of 3′-UTR. A: Schematic diagram of TPRT initiated by 3′-UTR; B: the conserved secondary structure of 3′-UTR (PDB: 8IBW), adapted from ref. [20] (Open Access); C: the core region of 3′-UTR, adapted from refs. [19,21] (Open Access); D: the conserved “pseudoknot-hinge-stem loop” structure of 3′-UTR from R2-A clade, adapted from ref. [21] (Open Access)

环2的密度更加完整、结构更为复杂, 展现了完整的假结(pseudoknot)结构[21,22]. Thawani等人[21]对包含R2*Tg*和R2*Pm*在内的五种R2-A分支的3′-UTR进行了序列比对, 发现它们均存在保守的“假结-铰链-茎环(pseudoknot-hinge-stem loop)”结构(图4D), 其中“假结”与R2*Bm*中的茎环2具有对应性, “茎环”与R2*Bm*中的茎环5具有对应性.

Wilkinson等人[19]以及Rodríguez-Vargas和Collins[53]的实验进一步证明了, 包含核心结构的截短体RNA即可模拟全长的3′-UTR, 起始TPRT过程(图4C). 这些研究共同阐明了3′-UTR起始TPRT的分子细节, 丰富了人们对3′-UTR结构与功能的认识. 然而, 目前已解析的结构均没有捕获到除核心结构以外的高分辨密度, 这可能与3′-UTR自身的结构柔性有关[19~22]. Thawani等人[21]观察到3′-UTR的“假结”与DNA底物的第二条链之间存在相互作用, 暗示DNA底物可能参与RNA的识别. 但是3′-UTR如何调控R2核酸内切酶活性的分子机制仍然需要更多结构生物学和单分子手段方面的研究.

2.3 5′-RNA的发现、结构与功能研究

2006年, Christensen等人[54]对大肠杆菌中纯化获得的R2*Bm*蛋白电泳样品进行银染时发现, 除R2*Bm*蛋白条带外, 还存在着一条对RNase A敏感的条带(~320 nt). 通过测序鉴定, 这段特异性结合R2*Bm*蛋白的RNA来源于R2转录本, 位于开放阅读框起始位点附

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

近，Christensen等人将这段RNA命名为5′-RNA(图5A)[54]. 眉纹天蚕蛾(*Samia Cynthia*, *Sc*)、赫拉克勒斯蛾(*Coscinocera Hercules*, *Ch*)、普罗米修斯天蚕蛾(*Callosamia promethea*, *Cpr*)、欧亚环纹天蚕蛾(*Saturnia pyri*, *Spy*)等多个物种来源的R2中也存在着5′-RNA的保守序列[55].

5′-RNA的偶然发现，使得人们重新评估R2*Bm*蛋白的活性. Christensen等人[54]对R2*Bm*蛋白样品进行了RNase A处理, 以消除内源5′-RNA对实验的影响, 分别探究了3′-UTR和5′-RNA对R2*Bm*活性的影响. 他们发现，单独的R2*Bm*蛋白具有对第一条链和第二条链的切割活性，5′-RNA的加入显著抑制了第一条链的切割活性，增强了第二条链的切割活性[54]. 为了完全消除5′-RNA在蛋白纯化过程的污染，Deng等人[20]对R2*Bm*

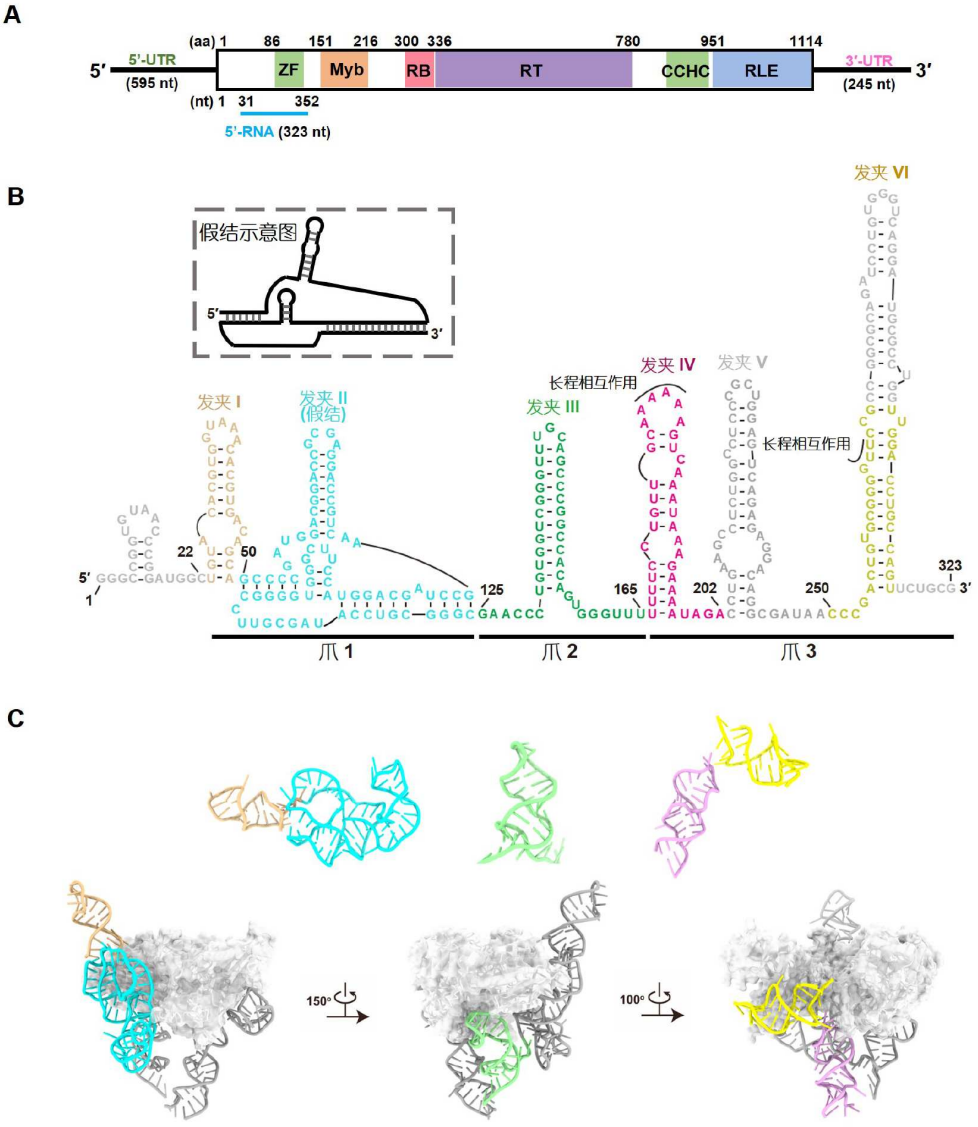


图 5　5′-RNA的结构. A: R2*Bm*蛋白质结构域及RNA示意图, 5′-RNA位于起始密码子(ATG中的A记为第1位碱基)下游第31位至第352位碱基; B: R2*Bm* 5′-RNA的二级结构和“假结”示意图. C: R2*Bm* 5′-RNA的三级结构(PDB: 8IBY). 图片修改自文献[20,55](开放获取)

Figure 5　The structure of 5′-RNA. A: Schematic diagram of protein domain and RNA of R2*Bm*. The 5′-RNA spans positions 31–352 (relative to the start codon defined as position 1); B: the secondary structure and pseudoknot of R2Bm 5′-RNA; C: the tertiary structure of R2*Bm* 5′-RNA (PDB: 8IBY). Adapted from refs. [20,55] (Open Access)

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

蛋白进行了密码子优化以破坏5′-RNA序列, 并使用离子交换柱去除核酸, 获得了单一、纯净的R2*Bm*蛋白样品. Deng等人[20]的实验表明, 单独的R2*Bm*蛋白质仅具有第一条链切割活性, 第二条链切割活性的激活依赖于5′-RNA. 因此, 5′-RNA在调控R2蛋白切割第二条链的过程中扮演了重要角色.

基于等能微阵列探针的二级结构预测[56]和核磁共振实验[57]显示, R2*Bm*的5′-RNA由六个发夹(hairpin)结构组成(发夹1~6), 其中发夹2呈现“假结”结构, 在R2*Sc*等多个物种中保守(图5B)[55]. 基于SHAPE技术的二级结构测定和冷冻电镜结构解析[20]表明, R2*Bm*的5′-RNA以复杂的“三爪(three-claw)”模式结合R2*Bm*蛋白(图5C). 其中, 包含发夹1和发夹2的爪1与R2*Bm*蛋白形成了连续的电荷相互作用界面, 占据了靶标DNA与R2*Bm*蛋白质结合的空间; 包含发夹4、发夹5和发夹6的爪3与R2*Bm*蛋白的RB结构域相互作用, 导致RLE结构域的构象发生改变, 并且包含爪3的5′-RNA截短体即可激活R2*Bm*蛋白对第二条链的切割活性.

结合结构分析和生化活性测定, Deng等人[20]对5′-RNA激活R2*Bm*蛋白切割第二条链的分子机制作出了推断: 5′-RNA的爪1与R2*Bm*蛋白的结合, 破坏了靶标DNA与R2*Bm*蛋白的作用界面, 为靶标DNA调整方向提供了空间, 5′-RNA的爪3与R2*Bm*蛋白的结合, 使得RLE结构域发生位移, 帮助捕获第二条链. 由第一条链切割状态转换至第二条链切割状态的分子细节仍然需要更多结构生物学数据的支撑.

值得注意的是, 尽管Kierzek等人[55]证明了5′-RNA序列在R2*Sc*, R2*Ch*, R2*Cpr*和R2*Spy*中的保守性, 但是目前仍然缺乏相关的结构和功能研究. 此外, 在R2*Tg*等R2中, 暂未发现与5′-RNA发挥类似功能的结构性RNA, R2*Tg*对第二条链的切割活性也不受5′-RNA的影响[21,23]. 因此, 5′-RNA的结构和功能在不同物种中的保守性仍然需要进一步的探究. 5′-RNA的存在与否以及其对转座过程的影响, 可能成为从机制层面对R2进行分类的新标准, 区别于R2-A至R2-D四个分支的分类体系, 将有助于R2逆转座子的差异和演化分析.

3 R2逆转座子工作机理的研究进展

R2精确靶向机制的阐明是工程化改造R2成为基因大片段插入工具的重要前提. 近年来, R2蛋白-RNA元件-rDNA复合物的高分辨结构解析, 显著推进了对R2工作机制的理解, 为R2工具的可控性设计和风险规避提供重要的理论支撑.

3.1 R2蛋白靶向28S rDNA的分子细节

在绝大部分物种中, R2均特异性地插入在28S rDNA的同一个位点. DNA印迹实验表明, R2蛋白对28S rDNA的靶向切割依赖于切割位点上游约40 bp长的序列[49]. Lee等人[40]比较了属于R2-A分支的R2*ZoAl*和赤拟谷盗(*Tribolium castaneum*, *Tc*)来源的R2*Tc*、属于R2-D分支的R2*Bm*和R2*Ds*, 系统性地分析了Myb结构域和不同ZF基序在DNA识别中的具体作用. 结果显示, 不管是在R2-A分支还是R2-D分支中, Myb结构域均具有很高的DNA结合亲和力并承担最主要的DNA识别作用, ZF1基序发挥了次要作用[40]. R2-A分支额外的ZF2和ZF3基序对DNA识别过程并不是必要的, 但它们能够促进R2工具在HEK293细胞中的基因插入效率[40]. 近期对R2*Pm*(R2-A分支)和R2*Bm*(R2-D分支)的冷冻电镜结构解析, 解释了不同分支的R2蛋白尽管在N端ZF基序的数量和种类上存在差异, 却仍能靶向同一位点的分子细节(图6A)[19~22,38]. 在R2*Pm*和R2*Bm*中, Myb结构域、ZF1基序和RT结构域通过静电相互作用和碱基特异性识别作用, 共同结合切割位点上游的DNA序列, 以捕获靶标DNA; 位于C端的CCHC及邻近基序和RLE结构域识别切割位点附近的DNA序列, 对双链DNA的解旋和靶标DNA进入RLE结构域活性口袋起至关重要的作用[19~22]. 相较R2*Bm*而言, R2*Pm*蛋白额外的ZF2及ZF3基序和Myb结构域及ZF1基序一起, 形成连续的相互作用界面, 进而保护靶标DNA, 发挥促进作用[21].

这些生物化学和结构生物学研究, 共同揭示了R2蛋白靶向28S rDNA的分子机理. 此外, 蛭形轮虫(*Adineta vaga*, *Av*)来源的R9*Av*属于R2-A分支, 但是靶向R9逆转座子所在的位点, 并产生了靶位点重复序列(target site duplication, TSD), 指示了R2存在非经典的插入模式[58]. 因此, R2在28S rDNA位点以外的脱靶频率测定和脱靶机制阐明仍然需要更多的研究.

3.2 R2蛋白识别模板RNA的分子细节

R2蛋白仅以包含3′-UTR序列的RNA作为模板起始TPRT过程, 这一特异性识别依赖位于RT结构域N端

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

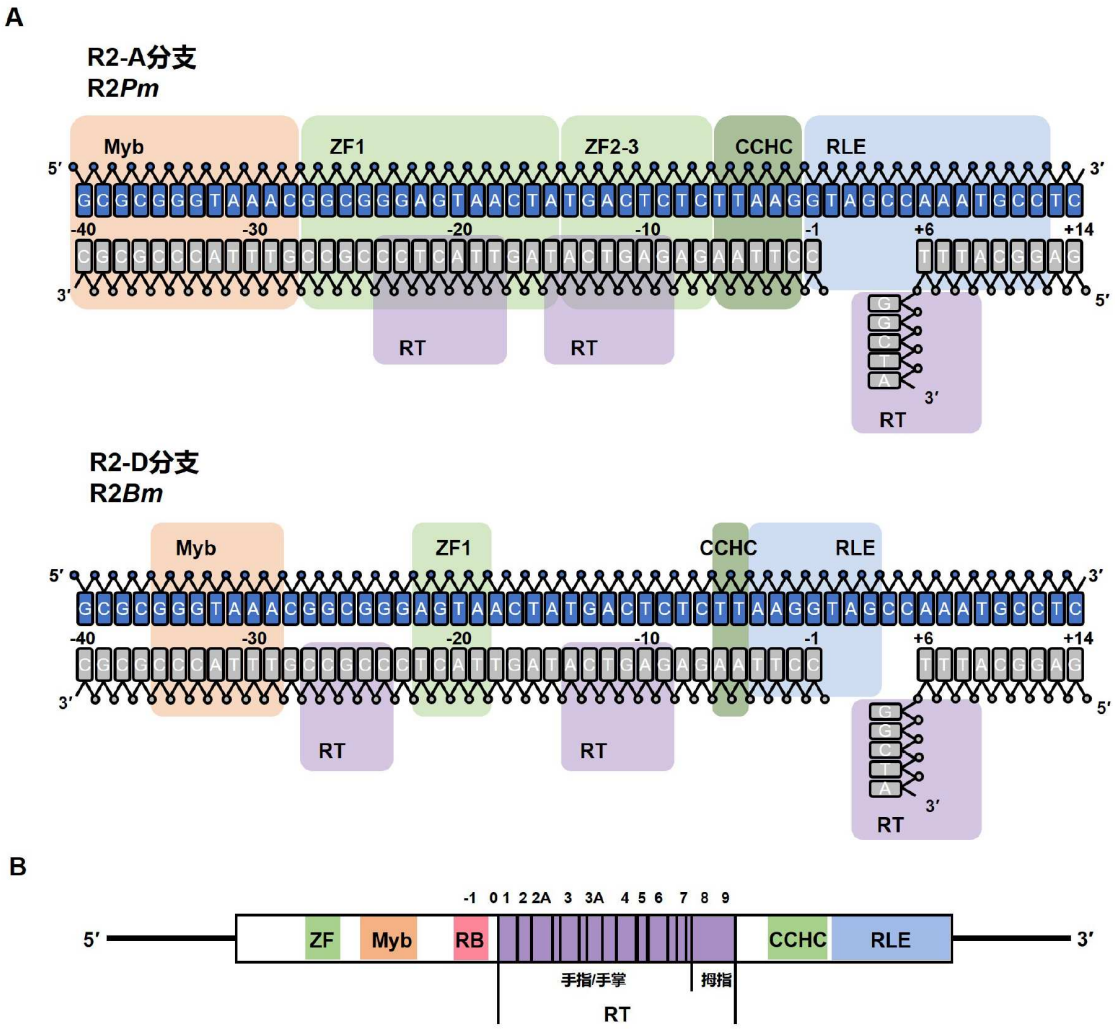


图 6　靶标DNA识别和RB结构域. A: R2*Pm*蛋白与R2*Bm*蛋白识别靶标的分子细节, 图片修改自文献[19,21](开放获取); B: RB结构域与RT结构域的排布

Figure 6　Target recognition and RB domain. A: The details of the target recognition by R2*Pm* protein and R2*Bm* protein, adapted from ref. [19,21] (Open Access); B: schematic diagram of RB and RT domain

的RB结构域[15,59]. 与经典的RT结构域相似, R2蛋白的RT结构域呈右手形状, 包含保守的手指(finger)、手掌(palm)和拇指(thumb)基序[60]. 根据L1, R1和R2等non-LTR逆转座子RT结构域的序列相似性, R2的RT结构域可以分为11个区域(图6B)[61]. 其中, R2独有的区域"–1"(即RB结构域)负责RNA的识别[19,20]. RB结构域的突变不会影响DNA合成活性, 但是显著降低了TPRT活性[59]. R2*Bm*, R2*Pm*和R2*Tg*的冷冻电镜结构均显示, RB结构域的关键氨基酸与RNA之间存在特异性的识别作用[19~22]. 在属于R2-D分支的R2*Bm*中, 目前的结构研究仅捕获了3′-UTR核心区域的茎环2和茎环5的高分辨密度, 其中, 茎环2主要由RB结构域的第307, 310和311位精氨酸识别[19,20]. 茎环2和茎环5之间连接处的A128由RB结构域的第314位酪氨酸、RT结构域的第732位亮氨酸和第733位赖氨酸共同识别; U129由RB结构域的第319位谷氨酸和第322位赖氨酸共同识别[19,20]. 在属于R2-A分支的R2*Tg*和R2*Pm*中, RB结构域(即NTE-1)同样发挥了识别3′-UTR的重要作用[21,22]. 此外, ZF2, ZF3, NTE和RT结构域还组成了连续的相互作用界面, 共同识别3′-UTR[21,22]. 以R2*Pm*为例, ZF3结构域的第33位精氨酸和第34位天冬酰胺对"假结"结构的识别至关重要; NTE-1结构域的第281位色氨酸和第

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

292位酪氨酸、RT结构域的第712位赖氨酸和第911位精氨酸共同识别了保守的“GAAAAG”铰链; NTE-2结构域的第264位赖氨酸和NTE-1结构域的第305位赖氨酸参与了“茎环”的识别[21].

3.3 R2转座机制探索的未来方向

R2的转座过程是一个多步骤反应, 包含模板RNA和R2蛋白的产生与组装、靶标DNA的识别与切割、双链DNA的合成与缺口修复. 尽管冷冻电镜结构解析揭示了R2蛋白识别模板RNA和靶标DNA的重要分子细节, R2蛋白如何精确、高效地完成双链DNA的合成并整合至基因组仍然有待探究.

2002年, Bibiłło和Eickbush[62]对R2的逆转录酶活性进行了细致的表征, 他们发现, R2可以在RNA模板之间进行多次、连续的跳跃, 并且“受体”RNA与“供体”RNA之间无需互补配对序列. 模板跳跃过程可能由RT结构域的寡核苷酸合成活性导致, 即RT结构域在到达“供体”RNA末端时, 可以随机添加非模板依赖的寡核苷酸序列, 从而与“受体”RNA的末端配对, 发生“端到端(end-to-end)”的跳跃[63]. Lentzsch等人[64]解析的第二类内含子成熟酶的晶体结构显示, 位于RT结构域上游的区域0对模板跳跃过程至关重要. 区域0的“PGPDG”序列在non-LTR逆转座子中高度保守(图6B), 关键氨基酸的突变将导致模板跳跃产物显著降低, 但是其中的分子细节仍然有待进一步研究[61]. RT的模板跳跃活性可能导致R2工具存在引入非目标序列的风险. 因此, 对R2蛋白模板跳跃机制的深入探究十分重要. 解析R2蛋白在模板跳跃状态下的结构, 寻找在非模板序列添加、稳定“供体”RNA与“受体”RNA末端和“端到端”之间跳跃等过程中起关键作用的氨基酸, 从而针对性设计蛋白突变体, 将有助于规避模板跳跃活性带来的风险.

此外, 人们对第二条链合成与缺口修复的机制仍然缺少足够的研究(图1C). 2007年, Kurzynska-Kokorniak等人[65]的研究发现, R2的RT结构域具有特殊的链置换活性, 这在DNA聚合酶中并不常见. 他们发现, 不具有RNase活性的R2可以利用DNA双链或者RNA:DNA异源双链中的单链DNA作为模板合成DNA, 这意味着R2具有链置换活性, 可以移除与模板DNA互补配对的单链DNA或RNA. R2特殊的链置换活性暗示了R2蛋白在转座过程中合成第二条链的可能性[65]. 然而, 目前尚未有明确的证据证明第二条链的合成由R2蛋白本身完成还是依赖细胞修复通路. 此外, 尚未有研究报道R2蛋白具有DNA连接酶活性, 如何修复合成的DNA双链与基因组之间的缺口仍然有待探究. 因此, 未来仍需探究R2依靠自身是否可以在细胞中完成完整的转座过程, 并尝试鉴定细胞修复通路中与R2活性相关的效应蛋白. 对第二条链合成和缺口修复的细致研究将帮助理解R2转座的完整过程, 提升R2工具的效率和可控性.

4 R2逆转座子的改造和工具开发现状

4.1 靶向28S rDNA的定向插入工具

2019年, Kuroki-Kami等人[24]使用青鳉鱼(*Oryzias latipes*, *Ol*)来源的R2*Ol*, 首次证明了R2*Ol*在细胞中的转座活性, 并成功在斑马鱼中实现了绿色荧光蛋白基因的敲入. 他们的初步尝试表明, 无脊椎动物来源的R2可以在脊椎动物细胞中工作, 并借助3′-UTR序列将荷载基因插入到28S rDNA中(图7A)[24]. 同年, Su等人[25]采用相似的策略, 证明了R2*Ol*同样可以在人类HEK293T细胞中实现基因敲入. 2024年, Chen等人[26]从Repbase数据库中挖掘了62类R2, 并系统性地表征了它们在人类HEK293T细胞中的转座活性. 其中, R2*Tg*展现出最佳的基因整合活性, 荷载基因的长度可达2.5 kb, 精氨酸突变筛选和逆转录酶结构域设计等策略均可进一步提升R2*Tg*的活性[26]. 同年, Zhang等人[27]的研究同样证明了R2*Tg*、白喉带鹀(*Zonotrichia albicollis*, *ZoAl*)来源和白喉鹀(*Tinamus guttatus*, *TiGu*)来源的R2均可作为基因大片段插入工具, 由此开发了以28S rDNA基因为安全位点, 定向整合外源基因的PRINT技术. 2025年, Edmonds等人[22]采用全RNA组分递送的形式, 证明了对RNA的5′端进行化学修饰可以进一步提升R2*Tg*工具的插入效率.

4.2 R2工具的重定向改造

28S rDNA作为物种间高度保守的多拷贝基因, 使得基于R2开发的基因大片段插入工具可以运用于斑马鱼、鼠和人等绝大多数真核细胞中[24,26]. 然而, 28S rDNA编码核糖体大亚基的重要组分28S rRNA, 对蛋白质翻译至关重要. 尽管对果蝇的研究表明, R2可以与28S rDNA共存, 并参与28S rDNA的拷贝数稳态调

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

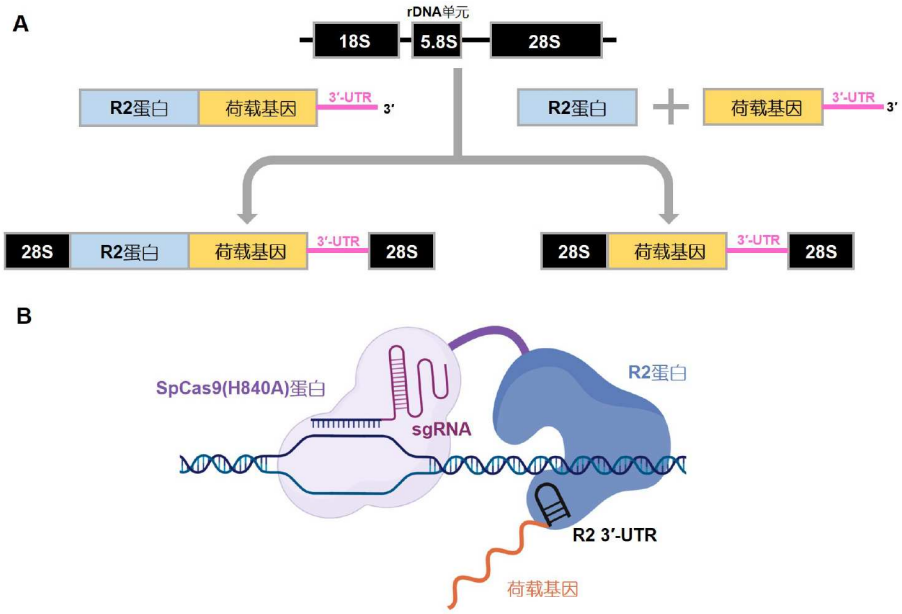


图 7　R2基因大片段插入技术. A: 利用R2向28S rDNA中插入荷载基因的原理; B: 融合Cas9蛋白实现R2的重定向改造
Figure 7　R2-based gene large-fragment insertion tools. A: The principle of inserting the payload gene into 28S rDNA using R2; B: retargeting payload insertion from 28S rDNA to desired loci via Cas9-R2 fusion design

控[66], 细胞中也存在多项调控机制共同维持28S rDNA的稳定性, 但是频繁的序列插入仍然可能带来极大的风险, 损坏细胞的正常功能[67]. Fefelova等人[68]发现, 相较于完整的rDNA基因, 存在R2插入的rDNA拷贝存在显著的HP1a和H3K9me3等抑制性表观修饰标志物的富集, 从而沉默R2的表达. 尽管PRINT等技术可以向28S rDNA基因中定向插入目的序列, 荷载基因是否可以成功表达仍然是无法预估的.

因此, R2工具的重定向改造是十分重要的. 2023年, Wilkinson等人[19]对R2*Bm*进行了融合蛋白改造, 初步实现了R2工具的重定向. 他们将R2*Bm*蛋白与HNH结构域失活的SpCas9^{H840A}融合, 成功在包含28S rDNA RUM序列(同Drr序列[20])和sgRNA靶向序列的嵌合DNA中检测到了TPRT产物[19]. 2025年, Fell等人[28]采用相似的策略, 进一步在人类HEK293T细胞中实现了R2*Tg*的重定向, 并基于此开发了插入序列长度可达12.7 kb的STITCHR技术(图7B). 然而, STITCHR等技术将R2蛋白与Cas9蛋白进行融合的设计, 使得完整蛋白的尺寸超过250 kD, 工具递送将成为新的挑战.

4.3　R2工具的未来改进方向

相较于PE[3], PASTE[4]和Amplification Editing[69]等技术, 基于R2开发的PRINT和STITCHR等技术具有多重优势: 全RNA成分的递送策略, 可以通过脂质纳米颗粒(lipid nanoparticle, LNP)实现高效率的递送, 同时避免了外源DNA的整合风险, 对遗传病治疗和CAR-T细胞疗法的开发具有重要意义; 以28S rDNA这类重复基因作为"安全港", R2工具可以在多个28S rDNA位点实现多基因、多拷贝的插入, 在合成生物学领域应用前景广阔; R2工具具备超过kb尺度的基因插入能力, 在现有技术基础上增加了可实现大片段基因插入的工具选择. 然而, 现有的R2工具仍然存在许多局限性. 如前所述, 以28S rDNA为定向位点的PRINT技术存在扰乱核糖体功能和荷载基因无法正常表达的潜在风险, 利用SpCas9^{H840A}实现重定向的STITCHR技术带来了递送难题. 由于3′-UTR序列是R2蛋白起始TPRT过程所必需的元件, 现有技术在实现荷载基因插入的同时, 均会引入3′-UTR序列, 尽管3′-UTR存在简化的空间, 然而目前基于R2的技术仍然无法实现无痕插入[19,20]. 值得一提的是, 5′端截短突变是L1, R1和R2等多种non-LTR逆转座子普遍存在的缺陷[70,71], R2逆转录酶在RNA模板之间的末端跳跃可能导致非目标序列的额外插入[63,72], 这些性质都使得荷载基因的精确、完整插入变得不可控. R2的RT结构域可以在低dNTP的浓

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

度下合成DNA，其错配频率与HIV-1的RT结构域相当，R2的保真性是否满足于工具化的要求仍然需要细致的表征[73]. 此外，以相似TPRT机制进行转座的L1被证明存在非核酸内切酶依赖的转座活性[74,75]，Tao等人[76]的数据进一步证明了L1可以在Cas9切割产生的DSB处进行广泛的转座. R2逆转座子是否存在类似的活性而带来脱靶风险, 仍然需要更加深入的探究.

5 总结与展望

自20世纪80年代在果蝇基因组中发现R2逆转座子以来，R2逆转座子的研究经历了三个关键阶段：在1990~2000年的早期研究中，人们主要通过基因组分析和生化活性表征的方式，认识到R2是一类以TPRT机制特异性插入到28S rDNA中的non-LTR逆转座子; 在2000~2015年前后，R2系统中5′-UTR，5′-RNA和3′-RNA这些RNA元件的发现和功能阐明，以及对RLE和RT结构域的细致研究, 帮助人们全面了解了R2逆转座子的特异性整合机制; 在2020年以后, 随着冷冻电镜等结构生物学技术的突破，R2蛋白识别模板RNA和靶标DNA的分子细节被阐明, 直接推动了R2系统在基因编辑工具开发中的应用转化. 目前, PRINT等技术可以实现在28S rDNA位点进行基因大片段的定向插入[26,27], STITCHR等技术利用融合Cas9的策略可以实现R2工具的重定向[19,28].

然而，人们对R2逆转座子的转座机制仍然存在知识盲区，R2的工具化依然面临多重挑战. 先前的研究已经基本阐明了R2逆转座子的TPRT机制，但是对于R2具有的模板跳跃和链置换等活性仍然缺乏细致的研究, 尤其是结构生物学方面的细节阐明; 第二条链的合成和缺口修复是否依赖于细胞修复通路，需要更多的实验数据证明；作为基因大片段插入工具，R2的脱靶效率、插入序列的突变频率等指标需要更加系统的表征；由R2介导的基因插入是否会引起表观遗传修饰的改变, 导致荷载基因的表达无法达到预期, 需要更多的检验. 截至目前，R2工具在使用灵活性、工作效率和安全性等方面的表现并未完全优于PE和PASTE等CRISPR衍生系统以及基于DNA转座子的CAST技术.

因此，未来的研究需要从机制和应用等多维度进行突破：在机制层面，运用结构生物学和单分子手段, 补全R2转座过程的细节，并对模板跳跃和链置换等的机制进行详尽的研究, 以规避R2工具可能存在的风险. 例如, 设计单分子荧光共振能量转移实验, 分别标记供体RNA和受体RNA，对模板跳跃反应中的动力学过程进行监测，并采用冷冻电镜技术捕获对应的构象状态, 从而深入理解其反应机理. 在应用层面, 筛选不同物种来源的R2，系统地表征R2工具的切割特异性和合成保真性，寻找表现最佳的R2系统，并运用定向进化、理性设计和人工智能辅助的蛋白质及核酸设计等手段, 改善现有的R2工具存在的位点单一、蛋白分子尺寸大、3′-UTR序列掺入等问题，发展更加安全、准确的基因大片段插入工具. 例如，目前对R2的工具化改造主要集中在R2-A分支和R2-D分支，广泛注释和挖掘R2及近缘的逆转座子，有希望获得与现有工具差异较大、性能更佳的候选系统. 此外，开展R2底物识别机制的细致研究，指导ZF基序和Myb结构域的重组和人工设计，有希望在不改变R2工具分子尺寸的前提下完成重定向改造. 随着对机制认识的深入和工具开发的成熟，R2逆转座子将成为认识和改造non-LTR逆转座子的典型范例, 启发更多分子工具的开发.

参考文献

1 Adli M. The CRISPR tool kit for genome editing and beyond. Nat Commun, 2018, 9: 1911

2 Rees H A, Liu D R. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet, 2018, 19: 770–788

3 Anzalone A V, Randolph P B, Davis J R, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 2019, 576: 149–157

4 Yarnall M T N, Ioannidi E I, Schmitt-Ulms C, et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nat Biotechnol, 2023, 41: 500–512

5 Anzalone A V, Koblan L W, Liu D R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol, 2020, 38: 824–844

6 Witte I P, Lampe G D, Eitzinger S, et al. Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase.

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

Science, 2025, 388: eadt5199
7 Lampe G D, King R T, Halpin-Healy T S, et al. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. Nat Biotechnol, 2024, 42: 87–98
8 Wells J N, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev Genet, 2020, 54: 539–561
9 Tang S, Sternberg S H. Genome editing with retroelements. Science, 2023, 382: 370–371
10 Di Nocera P P, Casari G. Related polypeptides are encoded by *Drosophila* F elements, I factors, and mammalian L1 sequences. Proc Natl Acad Sci USA, 1987, 84: 5843–5847
11 Burke W D, Calalang C C, Eickbush T H. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol Cell Biol, 1987, 7: 2221–2230
12 Finnegan D J. Eukaryotic transposable elements and genome evolution. Trends Genet, 1989, 5: 103–107
13 Bonnet A, Lesage P. Light and shadow on the mechanisms of integration site selection in yeast Ty retrotransposon families. Curr Genet, 2021, 67: 347–357
14 Luan D D, Korman M H, Jakubczak J L, et al. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell, 1993, 72: 595–605
15 Luan D D, Eickbush T H. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol, 1995, 15: 3882–3891
16 Cristofari G. Snapshots of genetic copy-and-paste machinery in action. Nature, 2024, 626: 40–42
17 Sultana T, van Essen D, Siol O, et al. The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection. Mol Cell, 2019, 74: 555–570.e7
18 Fujiwara H, Lambowitz A, Craig N. Site-specific non-LTR retrotransposons. Microbiol Spectr, 2015, 3: Mdna3
19 Wilkinson M E, Frangieh C J, Macrae R K, et al. Structure of the R2 non-LTR retrotransposon initiating target-primed reverse transcription. Science, 2023, 380: 301–308
20 Deng P, Tan S Q, Yang Q Y, et al. Structural RNA components supervise the sequential DNA cleavage in R2 retrotransposon. Cell, 2023, 186: 2865–2879.e20
21 Thawani A, Rodríguez-Vargas A, Van Treeck B, et al. Structures of vertebrate R2 retrotransposon complexes during target-primed reverse transcription and after second-strand nicking. Sci Adv, 2025, 11: eadu5533
22 Edmonds K H K, Wilkinson M E, Strebinger D, et al. Structure and biochemistry-guided engineering of an all-RNA system for DNA insertion with R2 retrotransposons. Nat Commun, 2025, 16: 6079
23 Thawani A, Collins K, Nogales E. Structural and biochemical studies of mobile retrotransposon proteins in action. Curr Opin Struct Biol, 2025, 92: 103053
24 Kuroki-Kami A, Nichuguti N, Yatabe H, et al. Targeted gene knockin in zebrafish using the 28S rDNA-specific non-LTR-retrotransposon R2Ol. Mobile DNA, 2019, 10: 23
25 Su Y, Nichuguti N, Kuroki-Kami A, et al. Sequence-specific retrotransposition of 28S rDNA-specific LINE R2Ol in human cells. RNA, 2019, 25: 1432–1438
26 Chen Y, Luo S, Hu Y, et al. All-RNA-mediated targeted gene integration in mammalian cells with rationally engineered R2 retrotransposons. Cell, 2024, 187: 4674–4689.e18
27 Zhang X, Van Treeck B, Horton C A, et al. Harnessing eukaryotic retroelement proteins for transgene insertion into human safe-harbor loci. Nat Biotechnol, 2025, 43: 42–51
28 Fell C W, Villiger L, Lim J, et al. Reprogramming site-specific retrotransposon activity to new DNA sites. Nature, 2025, 642: 1080–1089
29 Long E O, Dawid I B. Repeated genes in eukaryotes. Annu Rev Biochem, 1980, 49: 727–764
30 Dawid I B, Rebbert M L. Nucleotide sequences at the boundaries between gene and insertion regions in the rDNA of *Drosophila melanogaster*. Nucleic Acids Res, 1981, 9: 5011–5020
31 Eickbush T H, Robins B. Bombyx mori 28S ribosomal genes contain insertion elements similar to the type I and II elements of *Drosophila melanogaster*. EMBO J, 1985, 4: 2281–2285
32 Xiong Y, Eickbush T H. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol Cell Biol, 1988, 8: 114–123

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

33 Kojima K K. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol, 2003, 20: 351–361
34 Jakubczak J L, Burke W D, Eickbush T H. Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc Natl Acad Sci USA, 1991, 88: 3295–3299
35 Kojima K K, Seto Y, Fujiwara H, et al. The wide distribution and change of target specificity of R2 non-LTR retrotransposons in animals. PLoS One, 2016, 11: e0163496
36 Burke W D, Eickbush D G, Xiong Y, et al. Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. Mol Biol Evol, 1993, 10: 163–185
37 Burke W D, Malik H S, Jones J P, et al. The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol Biol Evol, 1999, 16: 502–511
38 Kojima K K, Fujiwara H. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. Mol Biol Evol, 2005, 22: 2157–2165
39 Thompson B K, Christensen S M. Independently derived targeting of 28S rDNA by A- and D-clade R2 retrotransposons. Mob Genet Elements, 2011, 1: 29–37
40 Lee R J, Horton C A, Van Treeck B, et al. Conserved and divergent DNA recognition specificities and functions of R2 retrotransposon N-terminal domains. Cell Rep, 2024, 43: 114239
41 Moss W N, Eickbush D G, Lopez M J, et al. The R2 retrotransposon RNA families. RNA Biol, 2011, 8: 714–718
42 Rosser J M, An W. L1 expression and regulation in humans and rodents. Front Biosci, 2012, E4: 2203–2225
43 Malik H S, Eickbush T H. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. Genetics, 2000, 154: 193–203
44 George J A, Eickbush T H. Conserved features at the 5′ end of *Drosophila R2* retrotransposable elements: implications for transcription and translation. Insect Mol Biol, 1999, 8: 3–10
45 Eickbush D G, Eickbush T H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. Mol Cell Biol, 2010, 30: 3142–3150
46 Eickbush D G, Burke W D, Eickbush T H, et al. Evolution of the R2 retrotransposon ribozyme and its self-cleavage site. PLoS One, 2013, 8: e66441
47 Yang J, Eickbush T H. RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. Mol Cell Biol, 1998, 18: 3455–3465
48 Christensen S M, Eickbush T H. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol Cell Biol, 2005, 25: 6617–6628
49 Christensen S, Eickbush T H. Footprint of the retrotransposon R2Bm protein on its target site before and after cleavage. J Mol Biol, 2004, 336: 1035–1045
50 Mathews D H, Banerjee A R, Luan D D, et al. Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. RNA, 1997, 3: 1–16
51 Ruschak A M, Mathews D H, Bibillo A, et al. Secondary structure models of the 3′ untranslated regions of diverse R2 RNAs. RNA, 2004, 10: 978–987
52 Duan S, Mathews D H, Turner D H. Interpreting oligonucleotide microarray data to determine RNA secondary structure: application to the 3′ end of *Bombyx mori* R2 RNA. Biochemistry, 2006, 45: 9819–9832
53 Rodríguez-Vargas A, Collins K. Distinct and overlapping RNA determinants for binding and target-primed reverse transcription by Bombyx mori R2 retrotransposon protein. Nucleic Acids Res, 2024, 52: 6571–6585
54 Christensen S M, Ye J, Eickbush T H. RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc Natl Acad Sci USA, 2006, 103: 17602–17607
55 Kierzek E, Christensen S M, Eickbush T H, et al. Secondary structures for 5′ regions of R2 retrotransposon RNAs reveal a novel conserved pseudoknot and regions that evolve under different constraints. J Mol Biol, 2009, 390: 428–442
56 Kierzek E, Kierzek R, Moss W N, et al. Isoenergetic penta- and hexanucleotide microarray probing and chemical mapping provide a secondary structure model for an RNA element orchestrating R2 retrotransposon protein function. Nucleic Acids Res, 2008, 36: 1770–1782
57 Hart J M, Kennedy S D, Mathews D H, et al. NMR-assisted prediction of RNA secondary structure: identification of a probable pseudoknot in the coding region of an R2 retrotransposon. J Am Chem Soc, 2008, 130: 10233–10239

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

58 Shivram H, Cawley D, Christensen S M. Targeting novel sites. Mob Genet Elements, 2011, 1: 169–178
59 Jamburuthugoda V K, Eickbush T H. Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. Nucleic Acids Res, 2014, 42: 8405–8415
60 Xiong Y, Eickbush T H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J, 1990, 9: 3353–3362
61 Malik H S, Burke W D, Eickbush T H. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol, 1999, 16: 793–805
62 Bibiłło A, Eickbush T H. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J Mol Biol, 2002, 316: 459–473
63 Bibillo A, Eickbush T H. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. J Biol Chem, 2004, 279: 14945–14953
64 Lentzsch A M, Stamos J L, Yao J, et al. Structural basis for template switching by a group II intron-encoded non-LTR-retroelement reverse transcriptase. J Biol Chem, 2021, 297: 100971
65 Kurzynska-Kokorniak A, Jamburuthugoda V K, Bibillo A, et al. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. J Mol Biol, 2007, 374: 322–333
66 Nelson J O, Slicko A, Yamashita Y M. The retrotransposon R2 maintains *Drosophila* ribosomal DNA repeats. Proc Natl Acad Sci USA, 2023, 120: e2221613120
67 Salim D, Gerton J L. Ribosomal DNA instability and genome adaptability. Chromosome Res, 2019, 27: 73–87
68 Fefelova E A, Pleshakova I M, Mikhaleva E A, et al. Impaired function of rDNA transcription initiation machinery leads to derepression of ribosomal genes with insertions of R2 retrotransposon. Nucleic Acids Res, 2022, 50: 867–884
69 Zhang R, He Z, Shi Y, et al. Amplification editing enables efficient and precise duplication of DNA from short sequence to megabase and chromosomal scale. Cell, 2024, 187: 3936–3952.e19
70 Pérez-González C E, Eickbush T H. Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans*. Genetics, 2001, 158: 1557–1567
71 Szak S T, Pickeral O K, Makalowski W, et al. Molecular archeology of L1 insertions in the human genome. Genome Biol, 2002, 3: research0052
72 Pimentel S C, Upton H E, Collins K. Separable structural requirements for cDNA synthesis, nontemplated extension, and template jumping by a non-LTR retroelement reverse transcriptase. J Biol Chem, 2022, 298: 101624
73 Jamburuthugoda V K, Eickbush T H. The reverse transcriptase encoded by the non-LTR retrotransposon R2 is as error-prone as that encoded by HIV-1. J Mol Biol, 2011, 407: 661–672
74 Sen S K, Huang C T, Han K, et al. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. Nucleic Acids Res, 2007, 35: 3741–3751
75 Morrish T A, Garcia-Perez J L, Stamato T D, et al. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature, 2007, 446: 208–212
76 Tao J, Wang Q, Mendez-Dorantes C, et al. Frequency and mechanisms of LINE-1 retrotransposon insertions at CRISPR/Cas9 sites. Nat Commun, 2022, 13: 3685

https://www.sciengine.com/doi/10.1360/SSV-2025-0147

Progress of research on the structure and function of RNA in R2 retrotransposon

TAN Shun-Qing[1,2,3,4*], YANG Yun[1,2,3,4], YANG Qi-Yu[1,2,3,4] & LIU Jun-Jie Gogo[1,2,3,4*]

[1] *School of Life Sciences, Tsinghua University, Beijing 100084, China*
[2] *Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing 100084, China*
[3] *Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China*
[4] *State Key Laboratory of Membrane Biology, Tsinghua University, Beijing 100084, China*
* *Corresponding authors, E-mail: tansq21@mails.tsinghua.edu.cn; junjiegogoliu@tsinghua.edu.cn*

The efficient and precise insertion of large DNA fragments remains a critical challenge in gene editing. While diverse technologies based on CRISPR-Cas systems enable base-level insertions, deletions, and substitutions, achieving high-precision integration of large DNA segments with high efficiency persists as a major bottleneck. Retrotransposons, natural genetic elements inherently capable of large-fragment insertion, offer a highly promising solution to this challenge. In recent years, significant advancements have been made in both the mechanistic understanding and the tool development of retrotransposons. Focusing on the R2 retrotransposon system, this review traces its developmental trajectory from the identification of key functional components through in-depth mechanistic elucidation to its engineering as a gene editing tool. The structure and function of the R2 RNA component are critically examined. This review consolidates innovative strategies for engineering R2 into an efficient large-fragment insertion tool.

RNA structure, gene editing, large-fragment insertion, non-LTR retrotransposon, R2 retrotransposon

doi: 10.1360/SSV-2025-0147

https://www.sciengine.com/doi/10.1360/SSV-2025-0147