# ITm活性转座子及其结构特征分析

潘春芳 汤定钦 周明兵\*

(浙江农林大学亚热带森林培育国家重点实验室培育基地,临安 311300)

摘要 ITm超家族是真核生物基因组中分布最广泛的DNA转座子家族之一,它们以DNA为媒介,通过"剪切-粘贴"机制在基因组中不断跳跃,引起基因组的重组与突变。随着对ITm转座子的深入研究,许多ITm转座子逐渐成为基因克隆、基因表达及其功能研究的重要工具。该文对活性ITm转座子作了较为全面的研究,并对其结构、拷贝数、分布以及转座特性进行了系统归纳,分析了活性ITm转座子螺旋-转角-螺旋(helix-turn-helix, HTH)结构,天冬氨酸-天冬氨酸-天冬氨酸/谷氨酸(Asp-Asp-Asp/Glu, DDD/E)催化结构域、Linker、核定位信号(nuclear localization sequence, NLS)及末端反向重复序列(terminal inverted repeats, TIRs)的序列特征。结果表明,自主转座活性的ITm转座子必须具备完整的转座酶及TIRs,上述结构及序列的突变均会不同程度的对转座子活性产生影响。这为活性ITm转座子的鉴定及功能分析奠定了重要基础,同时,也为人工调控ITm转座子转座活性提供了理论依据。

关键词 ITm转座子;转座;HTH结构域;DDD/E催化结构域;末端反向重复序列

# ITm Active Transposons and Analysis of Its Structural Characteristics

Pan Chunfang, Tang Dingqin, Zhou Mingbing\*

(The Nurturing Station for the State Key Laboratory of Subtropical Sil-viculture, Zhejiang A & F University, Lin'an 311300, China)

**Abstract** The *ITm* transposons superfamily is one of the most widespread transposons in eukaryotic organisms among DNA transposons. They jump in the host genome by the "cut-and-paste" mechanism via a DNA intermediate, which results in recombination and mutation of host genome. More and more *ITm* transposons have been developed to the genetics tool for gene cloning, gene expression and functional analysis. We comprehensively investigated characteristics of the active *ITm* transposons, including their structures, copy numbers, distributions and transposition characteristics. The features of HTH (helix-turn-helix) domain, DDD/E (Asp-Asp-Asp/Glu) catalytic domain, Linker, the NLS (nuclear localization sequence) and the TIRs (terminal inverted repeats) sequence were characterized. The results show that only the *ITm* transposons which have complete transposase structure and TIRs sequence could independent transpose in the host genome. The modifications of above mentioned domains of transposases and TIRs sequence would regulate the activity of transposons, but also for artificial regulation of transposition activity of these elements.

**Keywords** *ITm* transposons; transposition; HTH domain; DDD/E catalytic domain; TIRs

收稿日期: 2015-09-07 接受日期: 2015-11-09

浙江自然科学基金杰出青年项目(批准号:LR12C16001)和国家自然科学基金(批准号: 31270645、31470615)资助的课题

<sup>\*</sup>通讯作者。Tel: 0571-63731263, E-mail: zhoumingbing@zafu.edu.cn

Received: September 7, 2015 Accepted: November 9, 2015

This work was supported by the Natural Science Foundation for Outstanding Youth Project of Zhejiang (Grant No.LR12C16001) and the Natural Science Foundation of China (Grant No.31270645, 31470615)

<sup>\*</sup>Corresponding author. Tel: +86-571-63731263, E-mail: zhoumingbing@zafu.edu.cn

网络出版时间: 2016-01-13 14:09:04 URL: http://www.cnki.net/kcms/detail/31.2035.Q.20160113.1409.002.html

转座子(transposable elements, TEs)又称跳跃 因子,是指在生物细胞中能从同一条染色体的一个 位点转移到另一个位点,或者从一条染色体转移到 另一条染色体上的DNA序列<sup>[1]</sup>。转座子分为两类: 一是能够从细胞染色体的一个位点迁移(转座)到另 一个位点,引起生物基因组或基因的重组和变异, 加速生物多样性和进化速率,即DNA转座子,遵循 "剪切-粘贴"转座机制;二是能够在生物基因组中 大量扩增拷贝,是一类"自私基因",即RNA转座子, 遵循"复制--粘贴"机制<sup>[2]</sup>。

Tc1-Mariner超家族是真核生物基因组中分布 最广泛的DNA转座子家族之一,此家族转座子正被 成功用作基因工程和基因传递工具。它主要包括 三个家族的转座子:一是Tc1-like转座子,Tc1最初是 作为引起线虫多态性的一个重复序列而被发现,也 是线虫基因组中首次被鉴定的DNA转座子<sup>[3]</sup>,与Tc1 具有相似结构的转座子统称类Tc1转座子(Tc1-like elements, TLEs); 二是类mariner转座子, mariner最早 是在研究毛里塔尼亚果蝇(Drosophila mauristiana)白 眼基因的一个不稳定突变时发现的[4],此后,在其他 物种中也鉴定到这一类型转座子的存在,并且部分 具有转座活性,于是将这类转座元件称为类mariner 转座子(mariner-like elements, MLEs), 简称为MLE转 座子; 三是Pogo-like转座子, Pogo转座子是在黑腹果 蝇(Drosophila melanogaster)中发现的<sup>[5]</sup>,此后发现的与 其结构类似的转座子均称为类Pogo转座子(Pogo-like elements), 如Lemi1<sup>[6]</sup>、Tigger1<sup>[7]</sup>等。之后, Shao等<sup>[8]</sup>发现, TLEs和MLEs和来自细菌的IS630相似,因此,此超家 族又包括了IS630和其在原核生物和纤毛虫中的同源 序列,被命名为ITm(IS630-Tc1-mariner)超家族。因此, ITm超家族主要包括来自真核和原核基因组的四类 亚家族类群,包括Tc1-like转座子(TLEs)、mariner转 座子(MLEs)、Pogo-like转座子和IS630, 文本主要讨 论来自真核基因组的前三个类别。

# 1 ITm活性转座子

在动植物界中,虽然己鉴定出许多ITm转座子的存在,但绝大部分转座子含有终止子、移框和缺失现象,因而不具有转座活性<sup>[9]</sup>。Feschotte等<sup>[10]</sup>通过对水稻全基因组扫描,共鉴定出34个不同的mariner转座子,其中27个积累了插入/缺失、颠换/替换、移码等突变。根据文献报道,目前存在的活性ITm活性

转座子很有限,如TLEs中的Tc1、Tc3、SB(Sleeping Beauty)、FP(Frog Prince)等; MLEs 中的 Mos1、 Himarl、Hsmarl、Hsmar2、Famarl、Osmar5等; Pogo-like中的Pogo、Lemil、Tiggerl等。这些ITm 转座子结构相近,主要包括转座子末端颠倒重复序 列(terminal inverted repeats, TIRs)和中间的开放阅读 框(open reading frame, ORF)。(1)TIRs序列长度一般 为10~40 bp, 含有2个蛋白质结合元件。(2)ORF序列 长度一般为1 000~1 500 bp, 含有编码具有DNA结合 结构域和催化结构域的转座酶的基因。转座酶N-端 的DNA结合结构域有1~2个HTH结构域,能够识别 TIRs<sup>[11]</sup>; 催化结构域能够结合催化所必需的二价阳 离子Mg<sup>2+</sup>或Mn<sup>2+</sup>,催化结构域负责DNA的剪切和整 合反应[12]; DNA结合结构域与催化结构域之间有1个 Linker结构,负责调节转座酶空间结构<sup>[13]</sup>。在转座 酶中,还有1个与DNA结合结构域部分重叠的一分 或二分的核定位信号(nuclear localization sequence, NLS)<sup>[14]</sup>, 它能与入核载体相互作用, 使转座酶能被 运进细胞核。

#### 1.1 TLE活性转座子

TLE转座子除上述基本结构外,在2个HTH之间 还有1个调节TLE转座酶与DNA结合反应的类GRPR 结构<sup>[15]</sup>, C-端一般具有"DD34E"催化结构域。TLE转 座子是ITm超家族中种类最多,分布最广的一类<sup>[16]</sup>, 并广泛存在于鱼类、昆虫、线虫、真菌及其他高等 或低等生物中,但很多TLE转座子在垂直进化中已 失活。现己有不少学者将某些"化石"转座子重新激 活,如SB、Minos等。当然,TLE转座子中也存在着 天然活性的转座子,如Tc1、Impala等。

1.1.1 SB SB最早来源于鲑鱼, 是一个人工构建的TLE转座子。1997年, Ivics等<sup>[17]</sup>基于积累的系统发生数据, 利用生物信息学的手段, 对存在于鲑鱼科中的一个已失活的TLE转座系统进行了分子水平的重建, 唤醒了其转座活性。SB插入位点偏向于重复的TA, 如ATA TAT AT。在小鼠生殖细胞中SB 插入位点仅在相邻供体位点的3 Mb范围内, 整个潜在插入位点范围为5~15 Mb, 位点期待现象可能是TLEs的一个普遍特征<sup>[18]</sup>。

SB能在大多数脊椎动物中发生转座,且其转座过程依赖于宿主因子<sup>[19]</sup>。高迁移率族蛋白1(high mobility group box 1 protein, HMGB1)是影响SB转座的宿主因子之一<sup>[20]</sup>,它通过改变DNA的结构来调控

DNA复制和重组过程。SB转座频率在HMGBI基因敲 除的小鼠细胞急剧下降,而在野生型小鼠细胞中过度 表达会引起高频率转座。与之作用类似的还有阻止SB 转座子剪切后整合的BANF1(barrier to autointegration factor 1)蛋白[21]、参与SB转座酶表达反馈调控的 HMG2L1蛋白<sup>[22]</sup>以及修复SB转座后剪切位点的Ku70、 DNA-PKcs等<sup>[23]</sup>。不同修复因子对应的修复机制不同, 由此产生的足迹也不完全相同<sup>[24]</sup>。Yusa等<sup>[25]</sup>对SB局部 进行CpG甲基化处理能增强其转座效率。在小鼠红 白血病细胞内相同基因位点处通过染色体重组导入 CpG甲基化与非CpG甲基化的SB, CpG甲基化的SB转 座效率至少是非CpG甲基化的100倍。同时, CpG甲基 化也能提高从质粒转座到基因组上转座效率。此外, SB存在过量表达抑制(overproduction inhibition, OPI)现 象,在小鼠肝细胞内检测SB转座效率,发现当转座酶 与转座子的浓度比为1:10~1:25时,转座效率最高,然 而当转座酶浓度继续升高时,转座效率则会降低[26]。

Davis等<sup>[27]</sup>利用SB转座子系统从胎儿成纤维 细胞中生成4种人类诱导多能干细胞(induced pluripotent stem cells, iPSCs)。这类iPSCs均能表现出多 能干细胞的特性,包括体外分化为3个胚层衍生物 的能力。Hammer等<sup>[28]</sup>也利用该系统实现了对小鼠 前列腺移植体细胞的转基因。Been等<sup>[29]</sup>为了解组 织细胞肉瘤(histiocytic sarcoma, HS)的病理学,利 用SB作为诱变剂在小鼠体内进行了一个正向遗传 筛选来确定HS遗传动力。同时,他们通过分析转座 子插入位点鉴定得到28个与癌症相关的基因,包括 Nf1、Myc、Jak2、Pten等。对比其他转基因系统, SB具有定向插入的遗传特性,从而减少了插入突变 的风险。目前, SB已被广泛用作转基因和基因治疗 的工具<sup>[30]</sup>。

1.1.2 Tc1 Tc1最初是作为引起秀丽隐杆线虫 (Caenorhabditis elegans)多态性的一个重复序列而被 发现,也是线虫基因组中首次被鉴定的具天然活性 的TLE转座子<sup>[3]</sup>。Tc1保守拷贝数在大多数线虫菌株 染色体上约为30,而仅在两株线虫菌株上分散存在 着数百份拷贝,如Bergerac菌株中为400。

*Tc1*在某些野生型菌株的生殖细胞中会发生低频率转座,引起插入突变,并具有一定插入偏好性,如*Tc1*易插入在*gpa-2*基因中,插入位点为TA<sup>[31]</sup>。 *Tc1*在每个体细胞中的剪切频率是生殖细胞的1 000 倍<sup>[32]</sup>,其高频率剪切具有组织特异性,仅局限于体细 ·综述·

胞中。除1个拷贝外绝大多数Tc1无论在体细胞还是 生殖细胞都会产生不精确切除的转座足迹(5'端CA 和3'端GT)。在高拷贝数菌株如Bergerac中,Tc1的剪 切频率很高。尽管如此,它依然在某些特定的基因 组位点上比较稳定,这很可能使它作为良好的遗传 标记,Tc1已应用于线虫物理和遗传图谱的构建,如 Files等<sup>[33]</sup>已经将线虫肌动蛋白基因定位于连锁群V, Blumenthal等<sup>[34]</sup>也已经把卵黄蛋白基因定位于连锁 群X,他们都是利用在Bergerac基因中Tc1的存在所 引起的DNA多态性而定位的。

1.1.3 Minos Minos是由Franz等<sup>[35]</sup>在果蝇(Drosophila hydei)基因组发现的,它是一个人工构建而具有活性的转座子,拷贝数为5~30。在大多数情况下,Minos转座酶所催化的精确切除和整合没有侧翼DNA的参与。绝大多数Minos转座酶对转座子侧翼序列的剪切或者非常精确(即原转座子序列正常插入恢复),或者留下一个一端为6 bp的转座子足迹(5′端TAC GAG和3′端CTC GTA),包括靶位点重复序列(target site duplication, TSD)的2个TA碱基和末端其他4个碱基<sup>[35]</sup>。

Minos在多种昆虫胚胎生殖细胞染色体上形成 稳定的插入,该特性在小鼠体细胞和生殖细胞内也 有显现<sup>[35]</sup>。Minos在许多宿主细胞中可发生高效率 转座,尤其是在哺乳动物中,能导致基因组发生突变, 表明Minos是一种功能基因组分析的通用工具<sup>[36]</sup>。 Minos插入位点严格插入于TA中,在内含子与外显 子上有明显的插入偏差。它优先插入于内含子,不 会干扰靶基因的表达,但有时会导致附近外显子序 列缺失<sup>[37]</sup>。

1.1.4 其他TLE转座子 TLE类活性转座子除了上述应用较为广泛的3个转座子外,还有Impala、Tc3、FP和PPTN。

Impala是从枯萎病菌(Fusarium oxysporum)突 变体F24中分离得到的天然活性TLE转座子,其转座 足迹类似于Tc1,插入位点位于非编码区中的TA<sup>[38]</sup>。 2003年,Queiroz等<sup>[38]</sup>在青霉菌(Penicillium griseoroseum)中研究Impala转座,发现其转座足迹与插入位 点均与枯萎病菌一致,并为开发一种基于该转座子 的标签系统来检测与果胶酶生产及克隆相关基因提 供了依据。

Tc3是秀丽隐杆线虫(C. elegans)突变菌株TR679 中第二活跃的天然活性TLE转座子,拷贝数为15,插 入位点为TA,转座足迹一般为1~2个碱基(5′端C/CA 和3′端G/TG),但在某些转座事件中,不会留下任何 足迹<sup>[39]</sup>。

FP是从北方豹蛙(Rana pipiens)中分离得到经 人工重新构建而具有转座活性的TLE转座子<sup>[40]</sup>,插 入位点偏向于外显子中的TA,转座足迹一般为CTG 或CAG。FP在主要脊椎动物类群的代表细胞中都 具活性,其转座活性在某些细胞中高于SB。

PPTN,也称Passport,是在鲽(Pleuronectes plattessa)基因组中鉴定的TLE转座子<sup>[41]</sup>,是迄今所 发现的脊椎动物中唯一有天然转座活性的转座子, 拷贝数为200~300,插入位点偏向于外显子中的TA。 在人类基因组中,该转座子并不是随机插入,如在人 类的第1条染色体上插入事件为0,而在第6条染色 体上为6,而染色体1号长度是6号的2倍,这有利于 PPTN转座子成为一种有效研究功能基因学应用的 工具<sup>[41]</sup>。

### 1.2 MLE活性转座子

MLE转座子除基本结构外,转座酶上还有2个保守序列Try-Val-Pro-His-Glu-Leu(WVPHEL)和Tyr-Ser-Pro-Asp-Leu-Ala-Pro(YSPDLAP),两序列间隔150个氨基酸残基<sup>[42]</sup>,C-端具有"DD(34/37/39)D"催化结构域。MLE转座子存在的广泛性、宿主的多样性以及它们明显的不依赖宿主的转座机制,因此具有作为转化载体的潜能,并运用于基因治疗,成为治疗人类遗传缺陷症的有价值的工具,如Mos1、Himar1、Hsmar1等转座子。

1.2.1 Mos1 Mos1是从毛里塔尼亚果蝇(Drosophila mauritiana)中分离得到的具有天然活性的 MLE转座子,拷贝数为20~30,插入位点倾向于TA 丰富区,转座足迹一般为2~3个碱基(5'端CCA和3' 端GA/TGA)<sup>[43]</sup>。它广泛存在于多种生物中,如斑 马鱼(Barchydanio rerio)、利什曼原虫(Leishmania major)、鸡生殖细胞及埃及伊蚊(Aedes aegypti)等。

*Mos1*首先于体外实验中研究其各部分作用, 发现体外实验中它不需要形成发夹结构(YREKK), 只需要转座酶的存在即可发生转座,同样温度对 *Mos1*的影响也很明显<sup>[44]</sup>,温度偏高则转座活性低, 其主要是影响转座子切除后的重新插入。此外, *Mos1*通过转基因技术转入烟草中却检测不到转座 事件,而从烟草中提取的Mos1转座酶在体外能检 测到它与TIRs结合,剪切并整合事件,表明*Mos1*在 烟草中转座可能会受到烟草核蛋白的抑制[45]。

为了提高Mos1转座酶的活性,2009年,Germon 等<sup>[46]</sup>在Mos1上选择同一位点或相近位点进行单点 突变。用大肠杆菌筛选系统共获得5种高活性的突 变类型,并进行2个及2个以上的随机组合,使转座酶 活性提高至原来的200~800倍,甚至更高,但是过高 的转座活性会产生细胞毒性,导致细胞死亡。目前, 利用Mos1的插入突变特性已在线虫(C. elegans)基 因组上构建了突变体库,通过对特定基因进行插入 突变来分析该基因的功能<sup>[47]</sup>。此外,Mos1还成功应 用于结核分枝杆菌(Mycobacterium marinum)突变体 库的构建<sup>[48]</sup>。

1.2.2 *Himar1 Himar1是irritans*亚家族的成员之一,分离于角蝇(*Haematobia irritans*),拷贝数约为
17 000。它是通过PCR技术重新构建及体外突变而
得到的具有高活性的转座子<sup>[49]</sup>。

Himar1转座频率受转座子大小、转座所需的温度及时间、Mg<sup>2+</sup>浓度、转座酶浓度以及转座子载体的影响,如当转座酶浓度过高时反而会抑制Himar1转座子转座,从而降低转座频率;Himar1转座频率会随着转座子的增大呈现指数式下降等,Lampe等<sup>[50]</sup>和Akerley等<sup>[51]</sup>还将Himar1转座酶进行突变,在大肠杆菌检测到其活性提高至原来的10~50倍,并研究出了一种Himar1转座子系统,已经开发为人类呼吸道病原菌基因组突变的工具之一,从而使它成为一种有效研究基因功能的工具。Himar1插入位点位于弯曲DNA区域的TA,并对供体位点DNA侧翼序列产生一定的影响。2008年,Yang等<sup>[52]</sup>将Himar1进行改造,并成功使齿垢密螺旋体(Treponema denticola)突变体得以产生,这为促进齿垢密螺旋体生物学和发病机制研究奠定了基础。

1.2.3 *Hsmar1 Hsmar1*是第一个在人类基因 组中发现的*cecropia*亚家族转座子,拷贝数约为 200。2007年, Miskey等<sup>[53]</sup>用生物信息学方法重新 构建了该转座酶基因并命名为*Hsmar1-Ra*(Hsmar1reconstructed ancestral),它第一个在脊椎动物具有活 性的*MLE*转座子,该转座子系统可用于脊椎动物的 基因研究以及转座的动力学研究。

*Hsmar1*转座受pH值、转座酶浓度影响<sup>[54]</sup>。除 最适的pH值为8.5外,其他的pH值会降低剪切效率; 同时,高浓度的转座酶会由于转座酶OPI使得剪切效 率急速降低,最适转座酶浓度为25~50 nmol/L。此外, DNA负超螺旋结构有利于Hsmarl两端靠近并促进转座<sup>[55]</sup>。转座子供体和靶位点的拓扑结构也能影响 Hsmarl转座子的整合<sup>[56]</sup>。Hsmarl优先插入位点为 重复TA处。

1.2.4 其他MLE转座子 MLE类活性转座子除了 上述应用较为广泛的3个转座子外,还有来源于植 物基因组的Osmar5,来源于动物基因组的Hsmar2、 Mboumar-9和Famar1。

Osmar5是从水稻(Oryza sativa)中分离的具有天然活性的MLE转座子<sup>[10]</sup>。2006年, Yang等<sup>[57]</sup>在酿酒酵母(Saccharomyces cerevisiae)中对Osmar5的转座做了系统研究。Osmar5在5'端TA附近的CTC C处发生剪切并留下1~4 bp的转座足迹, 3'端则在GGA G出发生剪切并留下1~4 bp的转座足迹, Osmar5的发现为植物转座子基因标签开发提供新的工具。

Hsmar2是第二个在人类基因组中发现的 irritans亚家族转座子,拷贝数约为1000,它是经人 工构建而具有活性的转座子<sup>[58]</sup>。在不同类型的宿主 细胞中,Hsmar2的过量表达对宿主细胞产生不同影 响<sup>[59]</sup>。在HeLa细胞中,过量表达Hsmar2会对细胞产 生毒性,而在其他哺乳动物细胞中过量表达则不会。 Hsmar2也能在细菌和真核细胞中发生转座行为,但 转座足迹不是很明显。

Mboumar-9是从大头蚁(Messor bouvieri)卫星 DNA中发现的具天然活性的MLE转座子<sup>[60]</sup>。插 入位点为TA重复位点。在相同的转座条件下, Mboumar-9转座效率与Mos1<sup>[61]</sup>、Himar1相当。

Famar1来源于欧洲蠼螋(Forficula auricularia), 是利用PCR-ligation-PCR技术对1000个拷贝进行重 新构建而具活性MLE转座子,拷贝数约为44000<sup>[62]</sup>。 Barry等<sup>[63]</sup>鉴定了每个活性Famar1拷贝,发现它们的 转座活性范围很广,从Himar1活性的1/5到2倍。由 于不同亚家族的MLE转座子之间不会发生相互作 用<sup>[64]</sup>, Famar1可以作为一种有效的遗传工具来研究 该亚家族的其他转座子,拓宽了该类转座子的研究 领域。与之类似的转座子还有来自欧洲蜜蜂(Apis mellifera)的Ammar1。

# 1.3 Pogo-like活性转座子

Pogo-like转座子一般编码2个转座酶蛋白。有些转座子可通过某些方式将2个转座酶蛋白质形成 1个催化酶,如Lemil转座子可通过RNA剪切形成1 个转座酶<sup>[6]</sup>;而Tiggerl转座子编码的2个转座酶不能 进行RNA剪切形成1个转座酶<sup>[7]</sup>, C-端一般具有1个 "DD30~33D"催化结构域。相对*MLE、TLE*转座子 而言, *Pogo-like*转座子在生物基因组中分布有限、 种类较少, 典型的*Pogo-like*转座子也仅限于*Pogo、 Tiggers、Lemi1*等。

1.3.1 Pogo、Tigger1和Tigger2 Pogo首次是在 黑腹果蝇(D. melanogaster)的1个white-eosin突变体 中发现的转座子,绝大多数黑腹果蝇多为190 bp的 Pogo, 1.1~1.5 Kb的拷贝数为10~15,完整的Pogo全 长为2.1 Kb,但该转座子拷贝数较少,仅在该物种中 有分布<sup>[5]</sup>。该转座子两端的TIRs为21 bp,具有2个 ORF。这2个ORF可通过RNA剪切形成1个完整的转 座酶,该转座酶也具有DNA结合结构域和"DD30D" 催化结构域。Pogo转座酶DNA结合结构域上只含 有1个HTH结构,能特异性识别并结合在TIRs中的 12 bp<sup>[65]</sup>。

在人类基因组中还发现了与Pogo相关的其他 转座子如Tigger1、Tigger2<sup>[7]</sup>,它们的拷贝数分别为 3 000和1 000。这2个转座子也具有2个ORF,但这2 个ORF则不能像Pogo一样通过RNA剪切形成一个转 座酶。

1.3.2 Lemil Lemil是从哥伦比亚生态型拟南芥 (Arabidopsis columbia)中克隆得到的单拷贝Pogolike转座子<sup>[6]</sup>。它是第一个在植物基因组中发现的 Pogo家族转座子。Lemil编码的转座酶基因中有一 段内含子,并在氨基酸39位及385位存在终止子和移 码突变。Loot等<sup>[6]</sup>对该转座酶进行了重构,使Lemil 具有活性。

# 2 活性ITm转座子特征分析

为研究活性ITm超家族转座子的转座酶结构(图 1)及TIRs序列特征,我们从NCBI中下载上述转座子 序列(登录号见表1)。根据已有文献的报道及注解, 分别获得这些活性转座子HTH结构域、DDD/E催化 结构域、NLS、Linker以及TIRs序列,并通过这些序 列信息分析上述活性转座子的各结构域特征。

#### 2.1 活性ITm转座子结构分析

为了进一步分析活性ITm超家族转座子HTH结构域、DDD/E催化结构域、NLS、Linker及TIRs序列特征,我们利用NCBI上的注释信息将上述活性转座子的转座酶基因翻译成氨基酸序列,将保守域及TIRs序列用DNAMAN进行比对,同时,用Weblogo

转座子 Transposon	登录号 Accession number	类别 Class	来源物种 Species	天然或人 工构建 Natural or artificial	结构特征 Structure characteristics	应用 Applications	参考文献 References
SB	_	TLEs	Fish	Artificial	Two HTH, DD34E	Tools for gene therapy	[30]
Tc1	X01005	TLEs	C. elegans	Natural	Two HTH, DD34E	The actin gene was located in linkage group V, and the egg yolk protein gene was located in linkage group X using <i>Tc1</i>	[3,33-34]
Tc3	M77697	TLES	C. elegans	Natural	Two HTH, DD34E	_	[39]
Minos	X61695.1	TLEs	D. hydei	Artificial	Two HTH, DD34E	Tools for functional	[35-36]
Impala	AF282722	TLEs	F. oxysporum	Natural	Two HTH, DD34E	genomic analysis —	[38]
FP	AY261371	TLEs	R. pipiens	Artificial	Two HTH, DD34E	_	[40]
PPTN	AJ249085	TLES	P. plattessa	Natural	Two HTH, DD34E	Tools for functional genomic application	[41]
Mosl	X78906.1	MLEs	D. mauritiana	Natural	Two HTH, DD34D	Toos for construction of mutant library	[47-48]
Himar I	U11646	MLEs	H. irritans	Artificial	One HTH, DD34D	Developed as a tool for the human respiratory tract pathogen genome mutation	[12,49,51]
Hsmar1	U52077	MLEs	Human	Artificial	One HTH, DD34D	For the study of genes and dynamics of transposition of vertebrate	[53]
Hsmar2	U49974	MLEs	Human	Artificial	One HTH, DD34D	_	[58]
Osmar5	GQ379705	MLEs	O. sativa	Natural	Two HTH, DD39D	_	[10,57]
Mboumar-9	CAH03740	MLEs	M. bouvieri	Natural	Two HTH, DD34D	_	[60]
Famar1	AY155492	MLEs	F. auricularia	Artificial	One HTH, DD34D	_	[62]
Ammar1	AY155490	MLEs	A. mellifera	Artificial	One HTH, DD34D	_	[62]
Pogo	X59837	Pogo-like	D. melanogaster	Natural	One HTH, DD30D	_	[5]
Tigger 1	HSU49973	Pogo-like	Human	Natural	One HTH, DD33D	_	[7]
Tigger2	\$72489	Pogo-like	Human	Natural	One HTH, DD33D	—	[7]
Lemil	AC006161	Pogo-like	A. columbia	Artificial	One HTH, DD32D		[6]

	表1 活性 <i>ITm</i> 转座子结构
Table 1	The structure of active ITm transposon

在线软件作图,分析活性*ITm*超家族转座子转座酶 HTH结构域、DDD/E催化结构域、NLS、Linker及 TIRs序列的结构特征。

2.1.1 HTH结构域分析 由于活性ITm超家族转 座子中Pogo-like转座子数量有限,植物MLE转座 子Osmar自身的同源性较高,本文仅对TLE和动物

MLE转座酶的HTH结构域、催化结构域、Linker 和NLS(下同)进行分析。TLE和MLE转座酶N-端为 DNA结合结构域,而在DNA结合结构域中包含着重 要的HTH结构域。

在*TLE*中(图2), DNA结合结构域有2个HTH结构域(HTH1、HTH2), 分别位于N-端第20~50个氨基



★: 该结构在*ITm*不同亚家族中存在差异; 1: *TLE*中有2个HTH, 绝大多数*MLE*有1个, *Pogo-like*有1个; 2: *TLE*中存在GRPR-like结构, *NLE*、 *Pogo-like*不存在; 3: NLS在*ITm*不同亚家族转座酶上的位置稍有差异; 4: 仅*MLE*中存在Linker结构; 5: *TLE*催化结构域三联体结构为 "DD34E"、*TLE*为"DD34D"、*Pogo-like*为"DD30~33D"。

★: the differences in the structure of *ITm* in different subfamilies; 1: two HTH in *TLE*, one HTH in the vast majority of *MLE*, one HTH in *Pogo-like*; 2: GRPR-like structure in *TLE*, not in *MLE* and *Pogo-like*; 3: the positions of NLS in different *ITm* subfamily transposase are slight different; 4: Linker structure only in MLE; 5: the *TLE* catalytic domain is "DD34E", the *MLE* catalytic domain "DD34D", the one of *Pogo*-like element is "DD30~33D".



Weblogo图中每个位置氨基酸字母高度表示出现频率(下同)。 The height of the amino acid letters represents the frequency of occurrence of the amino acid in Weblogo (the same below).

图2 TLE转座酶HTH结构分析图 Fig.2 The HTH structure analysis of TLE transposase

酸,80~110个氨基酸之间。在MLE中(图3),DNA结 合结构域有1~2个HTH结构域,其分布与TLE类似。 结构及功能分析表明,该结构域是转座酶常见的结 构域,HTH1特异性识别并结合TIRs序列,使之形成 单末端复合体(single-end complex, SEC),是DNA转 座子发生转座的首要条件,HTH2则对HTH1结合起 加强作用<sup>[66]</sup>。在TLE中,2个HTH之间有1个GRPRlike结构,该结构负责调节TLE转座酶与DNA结合反 应<sup>[15]</sup>,在MLE中的Mos1中存在类似结构为GKPPK; 在TLE和MLE中均有CK-II磷酸化信号结构<sup>[67]</sup>,该结构调节转座酶翻译后修饰。

TLE中HTH结构域保守性较差(图2)。HTH1 结构域中I(异亮氨酸)、S(丝氨酸)、K(赖氨酸)、 L(亮氨酸)、T(苏氨酸)相对保守;HTH2结构域中 T、D(天冬氨酸)、L、V(缬氨酸)、G(甘氨酸)高度 同源,其中L在该结构域保守氨基酸中占3个。HTH 中某些氨基酸的突变会提高转座酶活性,如SB。 Mates等<sup>[68]</sup>通过定点突变方法对SB转座酶进行突变



图4 TLE转座酶DDE催化结构域分析图 Fig.4 The DDE catalytic domain analysis of TLE transposase

(HTH1:K33A, HTH2:R115H等其他4个氨基酸), 使其 活性提高至原来的100倍<sup>[30]</sup>, 并在人类Hela细胞和其 他脊椎动物细胞中对该转座酶进行了转座实验, 前 者SB100X的转基因效率达到了35%, 比野生型SB在 染色体上转座效率高至120倍。SB100X转座酶在斑 马鱼胚胎中瞬时转基因的效率达到了原来的8倍。

MLE中的HTH结构域保守性较好(图3)。除 Mos1和Mboumar-9外,其余转座酶仅有1个HTH结构。在HTH1结构域中,大多数氨基酸较为保守, 其中W(色氨酸)、F(苯丙氨酸)是保守氨基酸,这 2个氨基酸均带有芳香族R基团。HTH2结构域由 XTX<sub>[2]</sub>EIAEXLXVSX<sub>[7]</sub>LX(X为任意氨基酸,方括 号中为X的个数)组成, T、E(谷氨酸)、A(丙氨酸)、 L、S为保守氨基酸。在HTH结构域中, 某些氨基酸 的改变会使转座酶活性失活, 如Mos1中的L92A/R、 E94A、Q100E/R的突变<sup>[46]</sup>。Mos1中的R106对转座 酶DNA结合结构域的特异性识别起至关重要的作用, R106A的突变使转座酶丧失结合能力<sup>[11]</sup>。综上所述, HTH中的氨基酸可能对维持其"螺旋-转角-螺旋"结 构及促进与转座子结合具有重要作用<sup>[11]</sup>。

2.1.2 DDD/E催化结构域分析 TLE和MLE的C-端为转座酶催化结构域,在该结构中有一个转座酶 特有的"DDD/E"三联体结构。在TLE中,该三联体 结构表现为"DD34E"(图4), MLE中则为"DD34D"(图 5)。在该三联体结构中每个D及E附近均有一段保守的氨基酸序列。

TLE催化结构域比对结果如图4, D1附近的保守 序列为WX<sub>[2]</sub>VXWSDEXKX<sub>[2]</sub>LFG,相应位置的W、 S、D、E、K为保守氨基酸。在FP中,这几个位置的 氨基酸T152S以及R315C发生共同突变时,该转座酶 失活<sup>[40]</sup>。D2附近的保守序列是FQ<sub>[2]</sub>DNDPKHT,相 应位置的F、Q(谷氨酰胺)、D、N(天冬酰胺)、H(组 氨酸)、T为保守氨基酸。在SB中, D2前1个氨基酸 M243H突变时,转座酶活性提高至原来的2倍<sup>[30]</sup>。E 周围的保守序列为LXWPSQSPDLNPIENLW,在该 序列中绝大多数氨基酸为保守氨基酸。在Tc3中,该 区域中D254V突变后,活性稍稍降低<sup>[39]</sup>。D3的位置 也具有类似氨基酸。其中, SPDL这连续的4个氨基 酸在TLE和MLE中均存在。Liu等[13]对这个保守氨 基酸序列进行了结构上的研究,发现该保守序列与 WVPHEL序列相互作用,同时还通过A-P(脯氨酸)-S 这3个氨基酸与催化结构域中的D3这个活性位点相 连接。这意味着该保守序列在维持转座酶构象上 具有重要意义。此外, D1和D2之间有一段保守区域 1(conserved domain 1), G、M(甲硫氨酸)、W、F、L、 Y(酪氨酸)为保守氨基酸,其中,G较其他保守氨基 酸而言占的比例较多。从结构上看,它可能有利于 "DDE"三联体活性位点的聚集,并螯合某些二价金 属阳离子[69],从而使转座酶更好的发挥催化作用[13]。 在E后面还有一段保守区域2(conserved domain 2), 该区域中L、W、R(精氨酸)、V、G、T为保守氨基酸。 目前,该区域的功能还不了解,但在FP中,该区域的 某些氨基酸发生变化后该区域的某些氨基酸的突变

(如R315C),转座酶则失活<sup>[40]</sup>。

MLE催化结构域比对结果如图5, D1附近的保 守序列是FLXRIVTGDEKW,相应的位置F、L、R、 T、D、E、W为保守氨基酸。D2周围的保守序列 由RKX<sub>[2]</sub>VXFHHDNAPXHTSX<sub>[2]</sub>TR组成,V、H、 D、N、A为保守氨基酸。D3周围绝大多数为保守 氨基酸,尤其是HPPYSPDLAPSD。当D3区域中的 某些氨基酸突变时,转座酶活性会受影响,如Mos1 中, W268A-F-Y和D284E突变时活性大幅度降低甚 至失活<sup>[46]</sup>; Himarl中, H267R突变时活性提高至原 来的10倍<sup>[50]</sup>。D1和D2之间有一段保守区域2,绝大 多数氨基酸高度同源,其中,K、V、W、L、G、T、 Y、R、P为保守氨基酸。该保守区域也可能有利 于"DDD"三联体活性位点对二价金属离子的结合, 促进转座酶催化作用。与TLE类似,在D3后也有一 段保守区域3(conserved domain 3), F、G、I、L、R、 K、Y为保守氨基酸,该区域功能还不清楚,有待于 研究。上述结果表明,催化结构域中的氨基酸不仅 维持了转座酶发挥作用时的构象,同时也对转座子 的剪切及再插入具有重要作用[42]。

2.1.3 Linker分析 在MLE中(图3), HTH2结构域 后有一段高度保守的氨基酸序列(conserved domain 1),即WVPHEL,是HTH结构域与催化结构域的 Linker,也是MLE转座酶二聚体形成的接触面。Liu 等<sup>[13]</sup>对Mos1、Mboumar-9、Hsmar1及Famar1中的这 段保守序列进行了研究,并对Hsmar1的WVPHEL作 了突变,发现当W118R、V119S/T、E122R、L123S 突变后,Hsmar1转座酶活性分别提高至原来的30、7、 30及50倍,而当P120K、H121G突变后,转座酶构象



图5 MLE转座酶DDD催化结构域分析图 Fig.5 The DDE catalytic domain analysis of MLE transposase

发生转变,影响了HTH结构域与DNA的结合反应, 使其活性分别降低至原来的100倍及12.5倍。他们 还揭示了WVPHEL控制着催化功能的启动。而在 *TLE*中(图2),该保守序列并不是很明显,目前也没 有研究者对此进行研究。以上结果都暗示,Linker 在HTH结构和催化结构域之间的衔接起着桥梁作 用,HTH结构域和催化结构域并非独自行使功能, 它们均需要其他保守结构域的辅助才能更好地发 挥转座酶功能<sup>[13]</sup>。

2.1.4 NLS分析 转座酶属于核蛋白的一种,核蛋 白在动物、真菌和植物中高度保守。在一般情况 下,它在氨基酸序列中有一个特定的信号,称为核定 信号位序列。NLS是蛋白质的一个结构域,通常为一些短的氨基酸序列,它能与入核载体相互作用,使 蛋白质被转入细胞核。NLS由4~8个氨基酸组成,含 有P、K和R等。*ITm*转座酶在细胞核中发挥作用,转 座酶进入细胞核与*ITm*转座酶上NLS结构域关系密 切。虽然对于NLS序列目前没有完全确定,但有两 类NLS的氨基酸序列特征到目前为止有详细报道。一个是单一组分的NLS(K/R<sub>[4-6]</sub>),另一个是双组分的 NLS(K/R<sub>[2]</sub>X<sub>[10-12]</sub>K/R<sub>[3]</sub>),在某些情况下它们都有一些短的氨基酸残基,如K、R、H等<sup>[70]</sup>。

在TLE中,HTH2结构末端存在二分NLS(图2), 而在MLE中则位于HTH2前面(图3),该结构引导转 座酶进入细胞核。TLE和MLE的二分NLS均符合(K/ R<sub>[2]</sub>X<sub>[10-12]</sub>K/R<sub>[3]</sub>),K、R为保守氨基酸,这两个氨基酸 R基团均带正电。此外,在Mos1和Mboumar-9催化 结构域中(图5),D1后有一个单NLS,也符合(K/R<sub>[46]</sub>) 结构。这说明,NLS对于部分*ITm*转座子的转座活性 具有重要调控作用<sup>[70]</sup>。

#### 2.2 ITm活性转座子TIRs序列分析

所有的*ITm*超家族转座子都具有TIRs序列,但 不同类型的转座子TIRs在长度上有所差别。在*TLE* 中,大部分活性转座子TIRs长度在200~500 bp之间, 有些转座子的TIRs为27 bp左右(*PPTN*, Impala),而 *Tc4*的TIRs可达到774 bp。在*MLE*中,各个活性转 座子TIRs长度比较一致,为30 bp左右。*Pogo-like* 转座子长度与*MLE*较接近,为25 bp左右。虽然活 性*ITm*转座子TIRs序列长度表现各异,但它们都发 挥着共同的作用:与转座酶DNA结合结构域结合, 实现转座子的转座<sup>[11]</sup>。

在TIRs序列中,有2个转座酶结合位点。不同的

TIRs序列其结合位点长度并不一致,如SB的结合位 点位于DRs(direct repeats)区,长度为15 bp; FP的结 合位点位于DRs区,长度为21 bp; Osmar5的结合位 点也位于DRs区,长度为17 bp等。其中,Osmar5转 座子TIRs结合位点的突变会导致转座子和转座酶 的结合大幅度降低。这些特异性的结合位点对转 座子实现转座具有重要的意义。

此外, TIRs序列中的TA含量, 也会影响转座 子的转座效率,如Mos1、Mboumar-9。Mos1转座 子两端的TIRs有4个不匹配碱基,分别位于第31、 39、41、56个碱基处。在Mosl右端TIRs PEC晶体 结构中发现, TS上的A31及其互补碱基T26与DNA 大沟联系紧密,并被HTH1所识别;T39、A41及其 互补碱基则被2个HTH之间的Linker及DNA小沟 所识别,这3个碱基均不与Mosl转座酶DNA结合 结构域互作,而第4个裸露碱基A56则是嘌呤特异 性结合转座酶上的R183<sup>[69]</sup>。此外, TIRs上的碱基 G33、G22特异性结合HTH1中的K44、R48。研究 表明, Mos1中IRR的AT含量为64.3%, 而左端TIRs 中则为50%, Mos1转座酶对左端TIRs的亲和性是 右端TIRs的5~10倍<sup>[11,67]</sup>;当把左端TIRs替换成右端 TIRs时,转座酶与TIRs的亲和性至少降低50倍;当 把右端TIRs替换成左端TIRs时,亲和性则所提高26 倍,体外剪切实验表明其剪切频率达到9.1%<sup>[69,71]</sup>。 Mboumar-9上TIRs的3'端碱基G替换成A时,转座酶 与TIRs的亲和性提高了4倍<sup>[69]</sup>。

此外,两侧TIRs之间的序列也会影响转座效 率,如体内实验则发现影响*Mos1*转座的因素有两 侧TIRs之间的序列长度和结构<sup>[71]</sup>,序列长度越长则 转座活性越低,内部序列能否顺利折叠也是其影响 因素之一。TIRs之间序列的GC含量同样影响*Mos1* 的转座,但并不呈线性关系。这提示,TIRs的长度 及碱基组成对转座子转座效率有重要意义<sup>[11]</sup>。

# 3 结语与展望

随着人们在分子水平上对转座子结构和功能 认识的不断深化,利用转座子的特性,一些活性转 座子已被改造为转座子标签应用于基因治疗、基 因功能研究、突变体库构建及基于marker-free的 植物转基因技术等。

在动物中,多数转座子已被应用于动物转 基因,利用转座子作为动物转基因的工具有使用

安全、负载容量大和整合效率高等优点。目前 己被开发成动物转基因工具的主要代表有SB、 piggyBac转座子(PB)和Tol2转座子。SB可在大多数 脊椎动物细胞中将所携带的基因在动物细胞中稳 定整合及长期表达[27]。将来它很有可能替代目前 繁琐的基因敲除、RNA干扰技术和动物病毒转基 因系统;此外,SB在诱导多功能干细胞方面的应用 也越来越突出,利用该转座子的定向点特异性修复 能力与干细胞的多向分化性相结合,就可通过基因 修复来治疗多种疾病,且无需进行体外细胞培养。 Dupuy等<sup>[73]</sup>用鼠胚胎干细胞中的转座子作为诱变 鼠的遗传分析工具,将非同源片段插入到子代基因 位点进行基因修复[72]。这在功能基因组学研究及 基因治疗等领域发挥重要的作用。还有一些转座 子则被用于突变体库的构建,如Mos1;另有一些被 用于功能基因组分析,如Minos。

在植物中,多数转座子被应用于多种植物的基 因鉴定及突变体库的构建<sup>[74]</sup>,如水稻的Tos17、拟 南芥的EVD等。一些转座子则被用于植物遗传多 样性分析,如大麦的BARE-1;最重要的是,还有一 些转座子可被用于转基因的基因转移载体,如玉米 的Ac/Ds。虽然目前植物转基因技术有较为完整和 成熟的转基因体系,如农杆菌介导的转基因,但在 完整的转基因植株形成之后,标记基因就成为多 余,并且还带来潜在的环境和食品安全性问题,容 易引起人们对转基因生物安全性的疑虑。为避免 标记基因的不利影响,基于转座子开发的markerfree新型植物转基因技术有效的解决了这个问题。 Xuan等[75]利用Ac/Ds系统诱导基因组重排,在水稻 基因组中得到较高的删除、倒置、复制等基因组 重排效率。转座子在染色体中的转座可以使得选 择标记基因与目的基因摆脱连锁,最终得到无选择 标记基因的转基因植株,这在植物转基因方面将是 一个重大的突破。

本课题组致力于竹亚科MLE转座子研究, 并从毛竹(Phyllostachys edulis)中克隆到2个全 长的MLE转座子,通过酵母检测系统发现这2个 MLE(Phmar1、Phmar2)转座子具有转座活性, Phmar1和Phmar2转座足迹多为转座子两末端3~4 bp的序列,插入位点偏好于TA含量高的区域,倾 向插入基因附近或其内部等。目前,本课题组研 究的主要方向是通过生物信息学和蛋白质工程方 法改造和优化这2个转座酶,提高其催化转座的能力。同时,探讨人工调控转座子的插入位点的理论 基础,使其更好地应用于基因标签、基因转化等。

#### 参考文献 (References)

- Feschotte C, Jiang N, Wessler SR. Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 2002; 3(5): 329-41.
- 2 Evgen'ev MB. Mobile elements and genome evolution. J Mol Biol 2007; 41(2): 203-13.
- Bessereau JL. Transposons in *C. elegans*. Wormbook 2006; 18: 1-13.
- 4 Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 2007; 41: 331-68.
- 5 Tudor M, Lobocka M, Goodell M, Petit J, O'Hare K. The pogo transposable element family of *Drosophila melanogaster*. Mol Genet Genomics 1992; 232(1): 126-34.
- 6 Loot C, Santiago N, Sanz A, Casacuberta JM. The proteins encoded by the pogo-like Lemi1 element bind the TIRs and subterminal repeated motifs of the *Arabidopsis* emigrant MITE: Consequences for the transposition mechanism of MITEs. Nuleic Acids Res 2006; 34(18): 5238-46.
- 7 Smit FA, Riggs AD. Tiggers and other DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 1996; 93(4): 1443-8.

8

- Shao HG, Tu ZJ. Expanding the diversity of the IS630-Tc1mariner superfamily: Discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics 2001; 159(3): 1103-15.
- 9 Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 2007; 8(4): 272-85.
- 10 Feschotte C, Swamy L, Wessler SR. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with Stowaway MITEs. Genetics 2003; 163(2): 747-58.
- 11 Zhang L, Dawson A, Finnegana DJ. DNA-binding activity and subunit interaction of the mariner transposase. Nuleic Acids Res 2001; 29(17): 3566-75.
- 12 Hartl DL, Lohe AR, Lozovskaya ER. Modern thoughts on an ancient mariner: Function, evolution, regulation. Annu Rev Genet 1997; 31: 337-58.
- 13 Liu DX, Chalmers R. Hyperactive *mariner* transposons are created by mutations that disrupt allosterism and increase the rate of transposon end synapsis. Nuleic Acids Res 2014; 42(4): 1-9.
- 14 Ivies Z, Izsvak Z, Minter A, Hacker PB. Identification of functional domains and evolution of *Tc1-like* transposable elements. Proc Natl Acad Sci USA 1996; 93(10): 5008-13.
- 15 Benjamin B, Yves B, Corinne AG. Assembly of the *Tc1* and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica 2007; 130(2): 105-20.
- 16 Pocwierz-Kotus A, Burzynski A, Wenne R. Family of *Tc1-like* elements from fish genomes and horizontal transfer. Gene 2007; 390(1/2): 243-51.
- 17 Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular re-

construction of Sleeping Beauty, a *Tc1-like* transposon from fish, and its transposition in human cells. Cell 1997; 91(4): 501-10.

- 18 Miskey C, Izsvak Z, Kawakami K, Ivics Z. DNA transposons in vertebrate functional genomics. Cell Mol Life Sci 2005; 62(6): 629-41.
- 19 Izsvák Z, Ivics Z, Plasterk RH. Sleeping Beauty, a wide hostrange transposon vector for genetic transformation in vertebrates. J Mol Biol 2000; 302(1): 93-102.
- 20 Zayed H, Izsvák Z, Khare D, Heinemann U, Ivics Z. The DNAbending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nuleic Acids Res 2003; 31(9): 2313-22.
- 21 Lee MS, Craigie R. A previously unidentified host protein protects retroviral DNA from autointegration. Proc Natl Acad Sci USA 1998; 95(4): 1528-33.
- 22 Walisko O, Schorn A, Rolfs F, Devaraj A, Miskey C, Izsvák Z, et al. Transcriptional activities of the Sleeping Beauty transposon and shielding its genetic cargo with insulators. Mol Ther 2008; 16(2): 359-69.
- 23 Izsvak Z, Stuwe EE, Fiedler D, Katzer A, Jeggo PA, Ivics Z. Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. Mol Cell 2004; 13(2): 279-90.
- 24 Liu G, Cui Z, Aronovich EL, Whitley CB, Hackett PB. Excision of Sleeping Beauty transposons: Parameters and applications to gene therapy. J Gene Med 2004; 6(5): 574-83.
- 25 Yusa K, Takeda J, Horie K. Enhancement of Sleeping Beauty transposition by CpG methylation: Possible role of heterochromatin formation. J Mol Biol 2004; 24(9): 4004-18.
- 26 Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, Kay MA. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet 2000; 25(1): 35-41.
- 27 Davis RP, Nemes C, Varga E, Freund C, Kosmidis G, Gkatzis K, *et al.* Generation of induced pluripotent stem cells from human foetal fibroblasts using the Sleeping Beauty transposongene delivery system. Differentiation 2013; 86(1/2): 30-7.
- 28 Hammer KDP, Alsop JD, Buresh-Stiemke RA, Frantskevich K, Malinowski RL, Roethe LS, *et al.* A novel method for somatic transgenesis of the mouse prostate using the sleeping beauty transposon system. Prostate 2014; 74: 1-11.
- 29 Been RA, Linden MA, Hager CJ, DeCoursin KJ, Abrahante JE, Landman SR, *et al.* Genetic signature of histiocytic sarcoma revealed by a sleeping beauty transposon genetic screen in mice. PLoS One 2014; 9(5): 1-10.
- 30 Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, *et al.* Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet 2009; 41(6): 753-61.
- 31 Vos JC, de Baere I, Plasterk RH. Transposase is the only nematode protein required for *in vitro* transposition of *Tc1*. Genes Dev 1996; 10(6): 755-61.
- 32 Eide D, Anderson P. Insertion and excision of the *C.elegans* transposable element *Tc1*. Mol Cell Biol 1988; 8(2): 737-46.
- 33 Files JG, Carr S, Hirsh D. Actin gene family of *Caenorhabditis* elegans. J Mol Biol 1983; 164(3): 355-75.
- 34 Blumenthal T, Squire M, Kirttand S, Cane J, Donegan M, Spieth J, et al. Cloning of a yolk protein gene family from *Caenorhabdiris*

elegans. J Mol Biol 1984; 174(1): 1-18.

- 35 Pavlopoulos A, Oehler S, Kapetanaki MG, Savakis C. The DNA transposon *Minos* as a tool for transgenesis and functional genomic analysis in vertebrates and invertebrates. Genome Biol 2007; 8(Suppl 1): S2.
- 36 Kapetanaki MG, Loukeris TG, Livadaras I, Savakis C. High frequencies of Minos transposon mobilization are obtained in insects by using *in vitro* synthesized mRNA as a source of transposase. Nuleic Acids Res 2002; 30(15): 3333-40.
- 37 Metaxakis A, Oehler S, Klinakis A, Savakis C. Minos as a genetic and genomic tool in *Drosophila* melanogaster. Genetics 2005; 171(2): 571-81.
- 38 Queiroz M, Daboussi MJ. Impala, a transposon from Fusarium oxysporum, is active in the genome of *Penicillium griseoroseum*. FEMS Microbiol Lett 2003; 218(2): 317-21.
- 39 Luenen HV, Colloms SD, Plasterk RH. The mechanism of transposition of Tc3 in C. elegans. Cell 1994; 79(2): 293-301.
- 40 Miskey C, Izsvak Z, Plasterk RH, Ivics Z. The Frog Prince: A reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nuleic Acids Res 2003; 31(24): 6873-81.
- 41 Clark KJ, Carlson DF, Leaver MJ, Foster LK, Fahrenkrug SC. Passport, a native Tc1 transposon from flatfish, is functionally active in vertebrate cells. Nuleic Acids Res 2009; 37(4): 1239-47.
- 42 Prasad MD, Nurminsky DL, Nagaraju J. Characterization and molecular phylogenetic analysis of mariner elements from wild and domesticated species of silkmoths. Mol Phylogenet Evol 2002; 25(1): 210-7.
- Crenes G, Moundras C, Demattei MV, Bigot Y, Petit A, Renault S. Target site selection by the mariner-like element. Genetica 2009; 138(5): 509-17.
- 44 Sinzelle L, Jégot G, Brillet B, Rouleux-Bonnin F, Bigot Y, Augé-Gouillou C. Factors acting on Mos1 transposition efficiency. BMC Mol Biol 2008; 9(1): 106-19.
- 45 Thomas X, Hedhili S, Beuf L, Demattéi MV, Laparra H, Khong GN, *et al.* The mariner Mos1 transposase produced in tobacco is active *in vitro*. Genetica 2010; 138(5): 519-30.
- 46 Germon S, Bouchet N, Casteret S, Carpentier G, Adet J, Bigot Y, *et al.* Mariner Mos1 transposase optimization by rational mutagenesis. Genetica 2009; 137(3): 265-76.
- 47 Robert VJ, Bessereau JL. Manipulating the *Caenorhabditis* elegans genome using mariner transposons. Genetica 2010; 138(5): 541-9.
- 48 Picardeau M. Transposition of fly mariner elements into bacteria as a genetic tool for mutagenesis. Genetica 2010; 138: 551-8.
- 49 Lampe DJ, Churchill MEA, Robertson HM. A purified mariner transposase is sufficient to mediate transposition *in vitro*. EMBO J 1996; 15(19): 5470-9.
- 50 Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM. Hyperactive transposase mutants of the Himar1 mariner transposon. Proc Natl Acad Sci USA 1999; 96(20): 11428-33.
- 51 Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ. Systematic identification of essential genes by *in vitro* mariner mutagenesis. Proc Natl Acad Sci USA 1998; 95(15): 8927-32.
- 52 Yang Y, Stewart PE, Shi XG, Li CH. Development of a transposon mutagenesis system in the oral spirochete treponema

denticola. Appl Environ Microbiol 2008; 74(20): 6461-4.

- 53 Miskey C, Papp B, Mátés L, Sinzelle L, Keller H, Izsvák Z, *et al.* The ancient mariner sails again: Transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. Mol Cell Biol 2007; 27(12): 4589-600.
- 54 Bouuaert CC, Chalmers R. Transposition of the human Hsmar1 transposon: Rate-limiting steps and the importance of the flanking TA dinucleotide in second strand cleavage. Nuleic Acids Res 2009; 38(1): 190-202.
- 55 Bouuaert CC, Liu DX, Chalmers R. A simple topological filter in a eukaryotic transposon as a mechanism to suppress genome instability. Mol Cell Biol 2010; 31(2): 317-27.
- 56 Bouuaert CC, Chalmers R. Hsmarl Transposition is sensitive to the topology of the transposon donor and the target. PLoS One 2013; 8(1): 1-10.
- 57 Yang GJ, Weil CF, Wessler. SR. A rice *Tc1*/Mariner-like element transposes in *Yeast*. Plant Cell 2006; 18(10): 2469-78.
- 58 Robertson HM, Martos R. Molecular evolution of the second ancient human mariner transposon, Hsmar2, illustrates patterns of neutral evolution in the human genome lineage. Gene 1997; 205(1/2): 219-28.
- 59 Gil E, Bosch A, Lampe D, Lizcano JM, Perales JC, Danos O, et al. Functional characterization of the human mariner transposon Hsmar2. PLoS One 2013; 8(9): 1-12
- 60 Palomeque T, Carrillo JA, Muñoz-López M, Lorite P. Detection of a mariner-like element and a miniature inverted-repeat transposable element (MITE) associated with the heterochromatin fromants of the genus Messor and their possible involvement for satellite DNA evolution. Gene 2006; 371(2): 194-205.
- 61 Tosi LR, Beverley SM. Cis and trans factors affecting Mos1 *mariner* evolution and transposition *in vitro*, and its potential for functional genomics. Nuleic Acids Res 2000; 28(3): 784-90.
- 62 Lampe DJ, Witherspoon DJ, Soto-Adames FN, Robertson HM. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol Biol Evol 2003; 20(4): 554-62.
- 63 Barry EG, Witherspoon DJ, Lampe DJ. A bacterial genetic screen identifies functional coding sequences of the insect mariner transposable element famar1 amplified from the genome of the earwig, forficula auriculari. Genetics 2003; 166(2): 823-33.

- 64 Lampe DJ, Walden KKO, Robertson HM. Loss of transposase-DNA interaction may underlie the divergence of mariner-family transposable elements and the ability of more than one mariner to occupy the same genome. Mol Biol Evol 2001; 18(6): 954-61.
- 65 Wang HM, Hartswood E, Finnegan DJ. Pogo transposase contains a putative helix-turn-helix DNA binding domain that recognises a 12 bp sequence within the terminal inverted repeats. Nuleic Acids Res 1999; 27(2): 455-61.
- 66 Feschotte C, Osterlund MT, Peeler R, Wessler SR. DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. Nucleic Acids Research 2005; 33(7): 2153-65.
- Auge-Gouillou C, Hamelin MH, Demattei MV, Periquet G, Bigot
   Y. The ITR binding domain of the Mariner Mos-1 transposase.
   Mol Genet Genomics 2001; 265(1): 58-65.
- 68 Newman M, Lardelli M. A hyperactive Sleeping Beauty transposase enhances transgenesis in zebrafish embryos. BMC Res Notes 2010; 3(1): 282.
- 69 Trubitsyna M, Grey H, Houston DR, Finnegan DJ, Richardson JM. Structural basis for the inverted repeat preferences of *mariner* transposases. J Biol Chem 2015; 290(21): 13531-40.
- 70 Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. Classical nuclear localization signals: Definition, function, and interaction with importin alpha. J Biol Chem 2006; 282(8): 5101-5.
- 71 Auge-Gouillou C, Hamelin MH, Demattei MV, Periquet M, Bigot Y. The wild-type conformation of the Mos-1 inverted terminal repeats is suboptimal for transposition in bacteria. Mol Genet Genomics 2001; 265 (1): 51-7.
- 72 Dupuy AJ, Clark K, Carlson CM, Fritz S, Davidson AE, Markley KM, *et al.* Mammalian germ-line transgenesis by transposition. Proc Nat Acad Sci USA 2002; 99(7): 4495-9.
- 73 vanden Driessche T, Ivics Z, Izsvák Z, Chuah MKL. Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. Blood 2009; 114(8): 1461-8.
- Ovcharenko OO, Rudas VA, Kuchuk MV. Plant transposable elements and their application in genetics and biotechnology. Tsitol Genet 2006; 40(4): 68-80.
- 75 Xuan YH, Zhang J, Peterson T, Han CD. Ac/Ds-induced chromosomal rearrangements in rice genomes. Mob Genet Elements 2012; 2(2): 67-71.