



4种鲇基因组中 *Tc1/Mariner* 转座子的鉴定与演化

赵连鹏, 邵峰, 胡竞文, 张耀光, 彭作刚*

(西南大学生命科学学院, 淡水鱼类资源与生殖发育教育部重点实验室, 重庆 400715)

摘要: 为了探讨 *Tc1/Mariner* 转座子在4种鲇(沟鲇、黑斑原鲃、鲇和南方鲇)中的特征及在演化过程中的作用, 实验使用 *de novo* 预测和同源预测两种方法鉴定了4种鲇基因组中的 *Tc1/Mariner* 转座子并进行演化分析。通过对所获得序列的统计及分类, 结果显示, 4种鲇基因组中 *Tc1/Mariner* 转座子的含量分别为9.46%、2.70%、7.83%和8.54%, 并且分别被分为38、20、43和55个家族。基于转座酶催化区域的结构和系统发育分析的结果, 4种鲇的 *Tc1/Mariner* 转座子总共可以分为5个不同的亚类群: DD34E(*Tc1*)、DD35E、DD36E、DD37E和DD35D(*pogo*)。系统发育分析表明, 多数支系中鲇和南方鲇的转座子聚为姐妹群。插入时间分析表明, *Tc1/Mariner* 转座子的插入时间主要分布在0~5百万年前。大多数 *Tc1/Mariner* 转座子在0~1百万年前或2~3百万年前有一个突然的扩增。另外, 4种鲇的 *Tc1/Mariner* 转座子大部分插入到基因的内含子中。本实验通过对4种鲇中 *Tc1/Mariner* 转座子的鉴定、比较与演化分析, 为 *Tc1/Mariner* 转座子的深入研究奠定了基础。

关键词: 鲇; 基因组; 转座子; *Tc1/Mariner*; 演化分析

中图分类号: Q 785; S 917.4

文献标志码: A

转座元件(transposable element)或转座子(transposon)是一段可以在宿主染色体上移动的散在重复序列, 广泛分布在真核生物基因组中^[1]。20世纪40年代, 美国科学家McClintock^[2]研究玉米(*Zea mays*)的斑点突变时, 在九号染色体上发现一个可跳跃的断裂位点, 并命名为“Ac/Ds”元件, 之后这种元件被称为转座子。根据其转座机制, 转座子可以分为Class I 和Class II。Class I 元件也被称为RNA转座子, 主要通过“复制和粘贴”方式进行转座^[3]。Class II 元件也被称为DNA转座子, 主要通过“剪切和粘贴”方式进行转座。

基于TIR特征的不同, Class II 元件进一步分为不同的超家族。其中 *Tc1/Mariner* 超家族转座子在鱼类和两栖类基因组中的含量颇为丰富^[4]。*Tc1/Mariner* 转座子以秀丽隐杆线虫(*Caenorhabditis*

elegans)中的 *Tc1* 和马里塔亚果蝇(*Drosophila mauritiana*)中的 *mariner* 命名, 可通过它们编码的转座酶进行转座^[5-6]。*Tc1/Mariner* 转座子一般可分为两种类型, 一种具有完整的可读框, 编码一个340~360 aa的转座酶, 称为自主转座子; 一种为可读框缺失或移位, 编码非功能性的转座酶, 称为非自主转座子。自主 *Tc1/Mariner* 转座子全长1 300~2 400 bp, 其末端存在反向重复序列(TIRs), 侧翼是二核苷酸(TA)的靶位点重复序列(TSDs)^[7]。*Tc1/Mariner* 转座子编码的转座酶都具有DDE/D三联体催化基序。基于DDE/D基序的第二个和第三个催化残基之间的残基数, 目前 *Tc1/Mariner* 超家族大致可分为7个主要亚类群, 分别为DD34E(*Tc1*)、DD34D(*mariner*)、DD×D(*pogo*)、DD37D(*maT*)、DD37E、DD39D和DD41D(*rosa*)^[8-10]。

收稿日期: 2019-01-07 修回日期: 2019-06-13

资助项目: 国家自然科学基金(31572254)

通信作者: 彭作刚, E-mail: pzg@swu.edu.cn

随着测序技术的发展,越来越多物种的基因组信息被解析,这为运用生物信息学方法在基因组水平分析转座子提供了机会。目前已有研究者使用同源预测方法对小立碗藓(*Physcomitrella patens*)基因组中的2种Tc1-like转座子即PpTc1和PpTc2进行了鉴定和分析^[11]。Yuan等^[12]使用*de novo*预测方法对淡水型、海洋型和江海洄游型共计52种鱼重复组中Tc1/Mariner转座子的含量进行了统计,结果显示,真骨鱼类中淡水鱼类的Tc1/Mariner转座子比海水鱼类的更为丰富。Tc1/Mariner转座子的水平转移(horizontal transfer, HT)也是生物体之间一种常见的现象,尤其在昆虫中报道的较多^[13]。目前,关于沟鲈(*Ictalurus punctatus*)、黑斑原鲃(*Glyptosternon maculatum*)、鲃(*Silurus asotus*)和南方鲃(*S. meridionalis*)的Tc1/Mariner转座子家族层面的分类、活性、演化分析及与宿主基因功能关系的研究还不完善,只在沟鲈中有少数文献报道^[14]。本实验试图对4种鲃基因组中的Tc1/Mariner转座子进行鉴定与分类,并通过比较基因组学和演化分析来阐述此类转座子的结构、分类、扩增模式及分布等特征。

1 材料与方法

1.1 Tc1/Mariner转座子的鉴定及分类

从NCBI(<https://www.ncbi.nlm.nih.gov/genome/>)数据库下载沟鲈基因组序列,黑斑原鲃的基因组序列从GigaDB(<http://gigadb.org/>)数据库下载获得。鲃和南方鲃的基因组序列及注释信息由本实验室人员测序分析获得。在*de novo*预测中,使用RepeatModeler 1.0.7软件(<http://www.repeatmasker.org/RepeatModeler/>)对沟鲈、黑斑原鲃、鲃和南方鲃基因组中的转座子进行预测。在同源预测中,本实验以Rebase(<https://www.girinst.org/rebase/>)和NCBI数据库中的Tc1/Mariner转座酶序列为询问序列,通过同源预测的方法对4种鲃基因组分别进行TBLASTn(相似度>30%、覆盖度>50%、e值<10⁻⁵)比对。去除所获得数据中相似度较低、拷贝数少于3条以及gap较多的结果。使用CD-HIT软件^[15](-c 0.8, -n 5)对上述*de novo*预测和同源预测两种方法鉴定的序列进行去冗余。利用Perl脚本对去冗余后的序列进行all-blast-all聚类(遵从80-80-80规则^[3]),然后分别与对应的4种鲃基因组进行BLASTn比对,寻找每个家族的拷贝

<http://www.scxuebao.cn>

数(相似度>80%、覆盖度>80%、e值<10⁻⁵,其余设置为默认参数)。对每个家族的拷贝序列两端分别延长2 kb并使用MUSCLE 3.8.31软件^[16]对齐,手动鉴定转座子的边界。最后利用DAMBE 6.4.101软件^[17]和Ugene 1.31.0软件^[18]构建每个家族的一致序列。

1.2 Tc1/Mariner转座子结构分析及在基因组中的含量统计

使用TransDecoder 2.0.1软件预测转座子的开放阅读框(ORF)。用PHYRE 2(<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index/>)预测转座酶的空间结构。转座酶序列的核定位信号(NLS)使用PSORT II (<https://psort.hgc.jp/>)和PROSITE(<https://prosite.expasy.org/>)进行分析。最后,以本实验鉴定的4种鲃的一致序列为库,使用RepeatMasker 4.0.7软件(<http://www.repeatmasker.org/>)进行注释,并分别对4种鲃的Tc1/Mariner转座子含量进行统计。

1.3 系统发育分析

利用MUSCLE软件分别对4种鲃中具有完整转座酶(长度大于300 aa)的一致序列和数据库中已报道的一些Tc1/Mariner转座酶序列进行比对。然后选取比对结果的DD×D/E区域,利用MEGA 6.0软件^[9]中的最大似然法构建系统发育树,bootstrap值设定为1 000。

1.4 插入时间估计及插入位置分析

为了估计4种鲃中Tc1/Mariner转座子的插入时间,所有一致序列的拷贝用MUSCLE软件比对。利用DAMBE软件计算每个家族的一致序列与全长拷贝的分化度,然后利用公式 $T=K/2r$ ^[20]估算插入时间(T),其中 K 值为分化度, $r=2.5\times 10^{-8}$ 为青鳞(*Oryzias latipes*)中每年每个核苷酸位点的中性替换速率^[21]。另外,使用鲃和南方鲃的表达序列标签(EST)数据与具有完整转座酶的家族序列分别进行BLASTn比对,选取相似度和覆盖度均大于90%的结果用于转座活性分析。同时,为进一步了解Tc1/Mariner转座子在基因内或附近的分布情况及特点,使用Perl脚本统计Tc1/Mariner转座子的全长拷贝在基因外显子、内含子及其侧翼(均<2.5 kb)的位置信息。

2 结果

2.1 4种鲇中 *Tc1/Mariner* 转座子的鉴定、分类与特征分析

通过 *de novo* 预测和同源性搜索后的鉴定及一系列分类, 然后使用 RepeatMasker 4.0.7 软件进行统计, 得到 4 种鲇基因组中 *Tc1/Mariner* 转座子的含量分别为 9.46% (沟鲇)、2.70% (黑斑原鲃)、7.83% (鲇) 和 8.54% (南方鲇), 并且在沟鲇、黑斑原鲃、鲇和南方鲇中分别确定了 38、20、44 和 55 个 *Tc1/Mariner* 转座子家族序列。这些家族的类型、长度和拷贝数等特征详见表 1。从转座子拷贝数可以看出自主 *Tc1/Mariner* 家族序列中 *Tc1* 转座子占比较高, 并且大多数家族序列以 5'-CAGTG 开头, 以 3'-GTCAC 结尾; DD35E 转座子家族只存在于鲇和南方鲇中; DD36E 转座子家族在 4 种鲇都发现有 1 或 2 个, 且只在沟鲇中鉴定

到 DD35D (*pogo*) 转座子; 除沟鲇外, 其他 3 种鲇中都发现了 1 或 2 个 DD37E 转座子。非自主 *Tc1/Mariner* 转座子中, 沟鲇的长度大于 800 bp 的 *Tc1/Mariner* 转座子和小于 800 bp 的 *Tc1/Mariner* 转座子 (MITEs) 的数量相差不大 (拷贝数为 2 000 多个), 其余 3 种鲇中 MITEs 数量颇多, 如黑斑原鲃、鲇和南方鲇中 MITEs 家族拷贝数分别为 4 078、3 234、5 238 个。另外, 本研究把先前在沟鲇中鉴定的 *Tc1* 转座子 *Tip2* (DQ318917.1)^[22-23] 与沟鲇的 *Tc1/Mariner* 转座子家族序列进行比对, 发现 *Tip2* 转座子 (全长 1 009 bp) 与本研究所鉴定的一个长度为 1 651 bp 的非自主 *Tc1/Mariner* 转座子家族的一致序列存在约 88.34% 的相似度, 并且 TIR 基本相同, 表明它们属于同一家族的转座子。

2.2 4种鲇中 *Tc1/Mariner* 转座酶的结构分析

Tc1/Mariner 转座酶的 N 端为 DNA 结合区域, C 端为催化区域^[24]。通过 PHYRE 2、PSORT II 和

表 1 4种鲇中 *Tc1/Mariner* 转座子的特征

Tab. 1 Characteristics of *Tc1/Mariner* transposons in four catfishes

物种 species	类型 type	名称 name	家族数目 family number	长度范围/bp length range	TIR 长度范围/bp the length range of TIR	催化基序 catalytic motif	ORF 长度范围/aa the length range of ORF	拷贝数 copy number	TSD
沟鲇 <i>I. punctatus</i>	自主	<i>Tc1</i>	6	1 539~1 642	33~241	DD34E	319~364	2 847	
		<i>pogo</i>	1	1 873	18	DD35D	445	4	
		DD36E	2	1 220~1 221	28	DD36E	317~346	327	TA
	非自主	MITEs	5	385~558	14~50	-	-	2 036	
		<i>Tc1/Mariner</i>	24	809~1 655	10~264	-	-	2 244	
黑斑原鲃 <i>G. maculatum</i>	自主	<i>Tc1</i>	3	1 583~1 631	24~214	DD34E	338~365	77	
		DD36E	1	1 221	28	DD36E	346	24	
		DD37E	1	1 584	24	DD37E	337	10	TA
	非自主	MITEs	9	376~687	22~185	-	-	4 078	
		<i>Tc1/Mariner</i>	6	881~1 632	16~203	-	-	92	
鲇 <i>S. asotus</i>	自主	<i>Tc1</i>	15	1 595~1 671	29~257	DD34E	337~398	1 007	
		DD35E	2	1 643~2 393	23~27	DD35E	333~372	21	
		DD36E	2	1 220~1 246	25~30	DD36E	346~351	121	
		DD37E	1	1 563	25	DD37E	338	747	TA
	非自主	MITEs	10	203~671	13~96	-	-	3 234	
南方鲇 <i>S. meridionalis</i>	自主	<i>Tc1</i>	19	1 595~1 821	27~257	DD34E	329~398	1 064	
		DD35E	1	1 642	29	DD35E	333	20	
		DD36E	1	1 220	30	DD36E	346	114	
		DD37E	2	1 561~1 563	25~38	DD37E	338	154	TA
	非自主	MITEs	14	345~762	10~248	-	-	5 238	
	<i>Tc1/Mariner</i>	18	867~2 587	20~283	-	-	561		

注: 此表中的拷贝数指具有完整 TIR 的拷贝, 也叫做全长拷贝, - 代表无相应基序

Notes: The copy number in this table refers to a copy with a full TIR, also known as a full-length copy, - represents no relative motif exist

PROSTITUTE程序对4种鲈的自主*Tc1/Mariner*转座酶结构进行预测,结果显示,这4种鲈中*Tc1/Mariner*转座酶N端的DNA结合区域含有2个螺旋角螺旋(helix-turn-helix, HTH)基序,每个基序由3个 α -螺旋组成(图1,图2)。*Tc1*转座子中的第2个HTH基序内的第2个 α -螺旋具有部分重叠的核定位信号(nuclear localization sequence, NLS)。在这2个HTH基序之间有一个保守的类胃泌素释放肽受体(gastrin-releasing peptide receptor like, GRPR-like)基序,其中大多数*Tc1*转座子为GRPR残基。这与之之前6种新真骨鱼中的研究结果基本相符^[25]。同时还发现,DD35E、DD36E和DD37E类型转座子的HTH基序与*Tc1*转座子极其相似。另外,在*Tc1/Mariner*转座酶C端含有DDE(D)基序的催化区域,此催化区域主要负责DNA序列的剪切并参与转座反应。通过分析可以看到在第2个天冬氨酸和第3个谷氨酸/天冬氨酸之间有34~37个氨基酸。比较例外的是沟鲈中的一个*Tc1-3-*Ip**家族转座子,其催化基序为ND34E,其与之之前报道的大西洋鳕(*Gadus morhua*)中的*Tc1_2_Gm*转座子的基序相同^[25]。

2.3 系统发育分析

为了确定4种鲈之间及与一些已报道*Tc1/Mariner*转座子的系统发育关系,使用95条*Tc1/Mariner*转座酶序列进行了系统发育树的构建。结果发现,4种鲈中的*Tc1/Mariner*转座子可以分为5个不同的亚类群即*Tc1*、DD35E、DD36E、DD37E和*pogo*,其中大多数属于*Tc1*转座子(图3)。系统发育分析表明大多数支系中鲈和南方鲈的转座子聚为姐妹群;个别支系中沟鲈或黑斑原鲈的转座子聚为姐妹群。另外,*Tc1/Mariner-5-Sa*转座子似乎与其他支系关系较远,其与细菌的一种转座子IS630关系较近。值得注意的是,沟鲈中的*pogo*转座子与其他鱼类如红鳍东方鲀(*Takifugu rubripes*)、斑马鱼(*Danio rerio*)和斑点雀鲷(*Lepisosteus oculatus*)的*pogo*转座子单独聚为一个类群,部分DD35E转座子与*Tc1*转座子聚为一支。总的来说,4种鲈及其他物种转座子的系统发育关系与物种之间的亲缘关系基本一致。

2.4 4种鲈基因组中*Tc1/Mariner*转座子的插入时间估计

基于K2P模型(Kimura 2 parameter distances)^[26]来估算4种鲈中*Tc1/Mariner*转座子家族每个拷贝

的扩增情况,结果显示*Tc1/Mariner*转座子的插入大约发生在0~5百万年前(million years ago, Mya)(图4)。其中沟鲈、鲈和南方鲈的自主*Tc1/Mariner*转座子在0~1百万年前的拷贝数目较多,表明除黑斑原鲈外,其余3种鲈的自主*Tc1/Mariner*转座子在最近都有大规模扩增的趋势。另外,4种鲈中长度小于800 bp的非自主*Tc1/Mariner*转座子(MITEs)基本上都在2.5~3百万年前有一个快速扩增的趋势,尤其在黑斑原鲈中拷贝数颇多且呈现单峰状态。相比较而言,除黑斑原鲈外的其余3种鲈的MITEs比自主*Tc1/Mariner*转座子插入时间早。使用鲈和南方鲈的EST数据对其相应自主*Tc1/Mariner*转座子的活性及表达进行检测,发现鲈中*Tc1-3-Sa*、*Tc1-7-Sa*和*Tc1/Mariner-3-Sa*以及南方鲈中*Tc1-4-Sm*、*Tc1-9-Sm*和*Tc1/Mariner-3-Sm*等转座子家族匹配上大量的EST,表明这些转座子可能具有潜在的转座活性。

2.5 4种鲈基因组中*Tc1/Mariner*转座子的插入位置分析

非自主转座子MITEs由于序列较短,常分布于基因附近^[27]。沟鲈基因组中的*Tc1/Mariner*转座子的分布区域在染色体内和染色体之间没有显著差异^[14]。然而此类转座子在其他鲈基因内部及附近的位置信息尚未明确。通过对*Tc1/Mariner*转座子的插入位置进行分析,结果发现,此类转座子在4种鲈基因内部都有插入,并且大部分插入到基因的内含子中。鲈和南方鲈的基因侧翼位置转座子的分布较多,尤其MITEs转座子占绝大多数。另外,在外显子区域也有少数*Tc1/Mariner*转座子的存在,并且鲈和南方鲈中的数量要远多于其他两个物种。这些插入到外显子区域的基因包括鲈中与免疫相关的*IL7RA*基因和与嗅觉受体有关的*OR4S1*基因、南方鲈中与视网膜有关的*Rx1*基因等。

3 讨论

3.1 4种鲈中*Tc1/Mariner*转座子的含量、分类和相关性分析

*Tc1/Mariner*转座子广泛地分布于脊椎动物基因组中,尤其在鱼类和两栖类中颇为丰富^[7]。它们可以通过垂直转移(vertical transfer)和水平转移(horizontal transfer)的方式进行繁殖^[28]。正是这

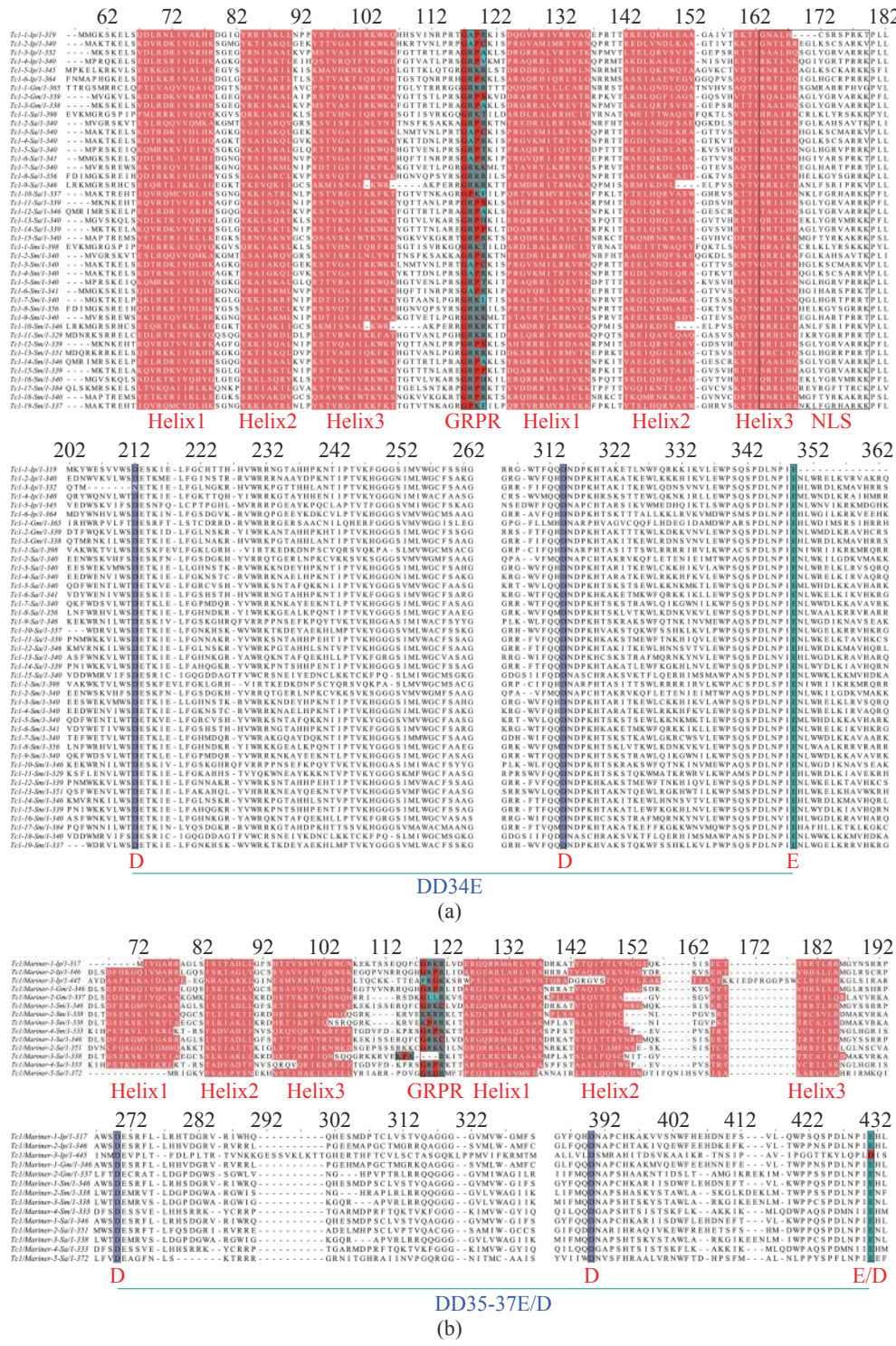


图 1 DD34E(a)和DD35-37E/D(b)Tc1/Mariner转座酶的多重比对结果

Ip. 沟鲈, Gm. 黑斑原鲈, Sa. 鲈, Sm. 南方鲈

Fig. 1 Multiple alignment of DD34E (a) and DD35-37E/D (b) Tc1/Mariner transposases

Ip. I. punctatus, Gm. G. maculatum, Sa. S. asotus, Sm. S. meridionalis

个特性, 使得转座子能够从垂直灭绝中逃脱, 并且演变得如此丰富^[29]。目前已经有研究者通过 *de novo* 预测对52种鱼类重复组中重复序列的含量

进行了估算^[12]。本实验通过 *de novo* 预测和同源性搜索两种方法对4种鲈的 *Tc1/Mariner* 转座子进行预测与分类, 并对其含量进行了估算。相比

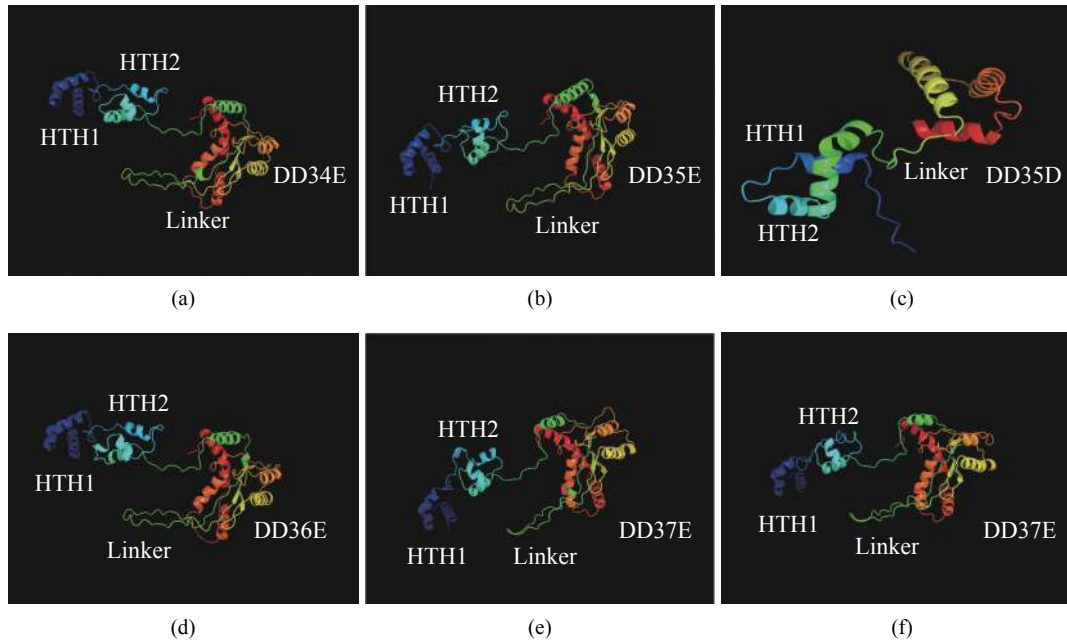


图2 4种鲇中自主 *Tc1/Mariner* 转座子的空间结构预测

Fig. 2 Predicted three-dimensional structures of *Tc1/Mariner* transposons in four catfishes

(a) *Tc1-10-Sa*, (b) *Tc1/Mariner-4-Sm*, (c) *Tc1/Mariner-3-Ip*, (d) *Tc1/Mariner-1-Gm*, (e) *Tc1/Mariner-3-Sa*, (f) *Tc1/Mariner-3-Sm*

黑斑原鲃(662.34 Mb, 2.70%)以及先前报道的绿河鲃(*Tetraodon nigroviridis*)(342.40 Mb, 0.51%)、红鳍东方鲃(391.49 Mb, 0.87%)、三刺鱼(*Gasterosteus aculeatus*)(446.61 Mb, 0.80%)、青鲃(734.06 Mb, 2.91%)、大西洋鲃(824.31 Mb, 0.25%)和尼罗罗非鱼(*Oreochromis niloticus*)(1 005.68 Mb, 5.09%)6种新真骨鱼(neoteleost)^[25], 沟鲃(783.19 Mb, 9.46%)、鲇(767.18 Mb, 7.83%)和南方鲃(762.47 Mb, 8.54%)的*Tc1/Mariner*转座子含量相对较高。对于脊椎动物而言, 转座子的含量与基因组大小一般呈正相关关系^[30]。本实验结果显示, *Tc1/Mariner*转座子的含量与4种鲃和6种新真骨鱼的基因组大小相关性并不显著($r = 0.57$, $P > 0.05$)。通过对自主*Tc1/Mariner*转座子的转座酶分析, 发现4种鲃中绝大多数为*Tc1*转座子。先前有研究者对6种新真骨鱼的*Tc1/Mariner*转座子的鉴定也发现很多*Tc1*转座子, 并且在这6种新真骨鱼中都发现了一个DD35D(*pogo*)转座子^[25]。然而4种鲃只有沟鲃中发现一个*pogo*转座子家族。另外, 张化浩^[31]在对家蚕(*Bombyx mori*)基因组分析时发现大多数*Tc1/Mariner*转座子属于*maT*类型。Bouallégué等^[32]在对豌豆长管蚜(*Acyrtosiphon pisum*)等蚜虫基因组中的转座子进行研究时, 发现了许多*mariner*、DD40D和*rosa*转座子。本实验

<http://www.sxuebao.cn>

在对4种鲃分析时, 并未鉴定到这些类型的转座子。由此可见, 鱼类*Tc1/Mariner*转座子的类型与昆虫存在较明显的差异。

3.2 4种鲃中 *Tc1/Mariner* 转座子的活性和扩增模式

*Tc1/Mariner*转座子虽然分布很广, 但很多都逐渐失去活性, 这可能是由于大多数编码转座酶的核苷酸序列受到缺失、移码和替代所造成^[33]。据报道, 当生物体在自然环境中受到压力时, 一些转座子可能被激活^[34]。目前在硬骨鱼类中发现了一些具有活性的*Tc1/Mariner*转座子, 如欧洲鲈(*Pleuronectes platessa*)中的*Passport*^[35]、意大利鲟(*Acipenser naccarii*)中的*Tana1*^[36]、鲢(*Hypophthalmichthys molitrix*)中的*Thm3*^[37]转座子等。本实验通过比对分析发现, 除黑斑原鲃不明显外, 其余3种鲃的自主*Tc1/Mariner*转座子在约0~1百万年前有一个突然的扩增, 如鲃中的*Tc1-3-Sa*以及南方鲃中的*Tc1-4-Sm*转座子等。从这些转座子的高拷贝数以及匹配上大量EST数据的特点可推测部分自主*Tc1/Mariner*转座子近期可能存在潜在的活性, 但需后续实验进一步验证。MITEs转座子可能起源于共同家族的自主转座子, 这些没有编码能力的MITEs, 需借助自主转座子的转座酶才能进行转座^[38]。本研究使用*de novo*预测

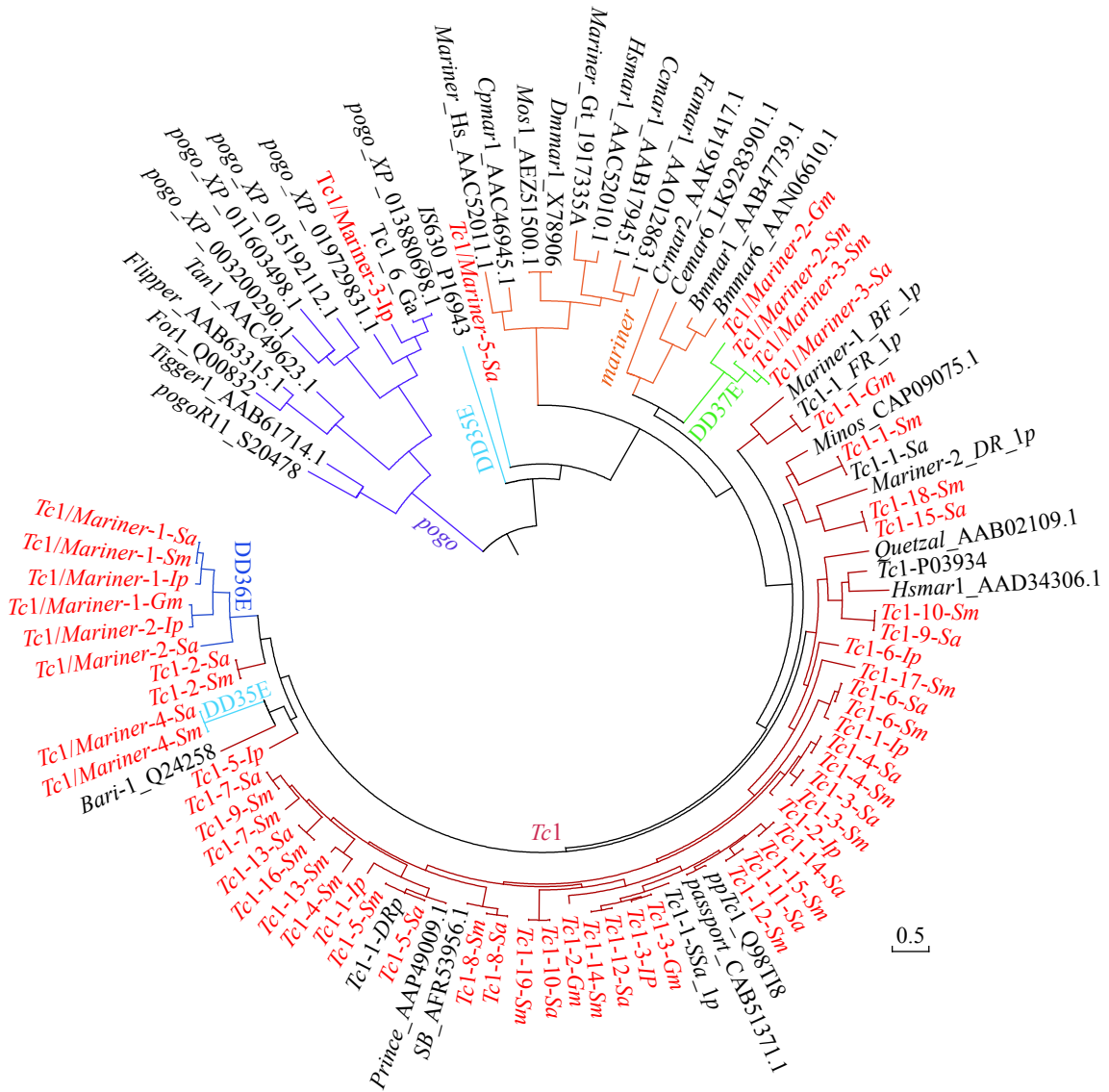


图 3 4种鲇及部分已报道过的Tc1/Mariner转座子的系统发育分析

分支颜色表示不同的Tc1/mariner转座子类群, 深红色代表DD34E(Tc1)、橙色代表DD34D(mariner)、紫色代表DDxD(pogo)、浅蓝色代表DD35E、蓝色代表DD36E、绿色代表DD37E、红色的分类名表示4种鲇的家族名称

Fig. 3 Phylogenetic analysis of Tc1/Mariner transposons in four catfishes and previously reported families

Clade colors denotes the different subfamilies of Tc1/Mariner transposons, dark red represents DD34E(Tc1), orange represents DD34D(mariner), purple represents DDxD(pogo), light blue represents DD35E, blue represents DD36E, and green represents DD37E; taxon name in red indicates family names of the four catfishes

和同源性搜索方法在4种鲇中并未检测到MITEs家族相应的自主Tc1/Mariner转座子, 其或许存在于4种鲇或其他物种中, 只是由于鉴定方法或基因组质量问题而暂未找到。

另外, 也有研究者提出转座子的“生命周期”假说^[39]。该假说认为, 新的活性转座子元件插入到宿主基因组后会有快速扩增的时期, 从而使该元件的拷贝数和多样性增加。另外由于突变或负选择的积累导致转座子垂直失活(vertical inactivation), 随后转座子拷贝数会达

到一个动态平衡, 最终非活性拷贝不再扩增直至缓慢消失。对Tc1/Mariner转座子插入时间的分析结果显示, 4种鲇的MITEs转座子大约在3~5百万年前出现, 紧接着主要在2~3百万年前有一个爆发式的扩增, 最后基本在0~1百万年前开始衰退。这种现象似乎符合转座子的“生命周期”假说。总之, 对4种鲇的Tc1/Mariner转座子插入时间和活性分析发现, MITEs的繁殖趋势基本符合转座子的“生命周期”假说, 并且近期经历着衰退。

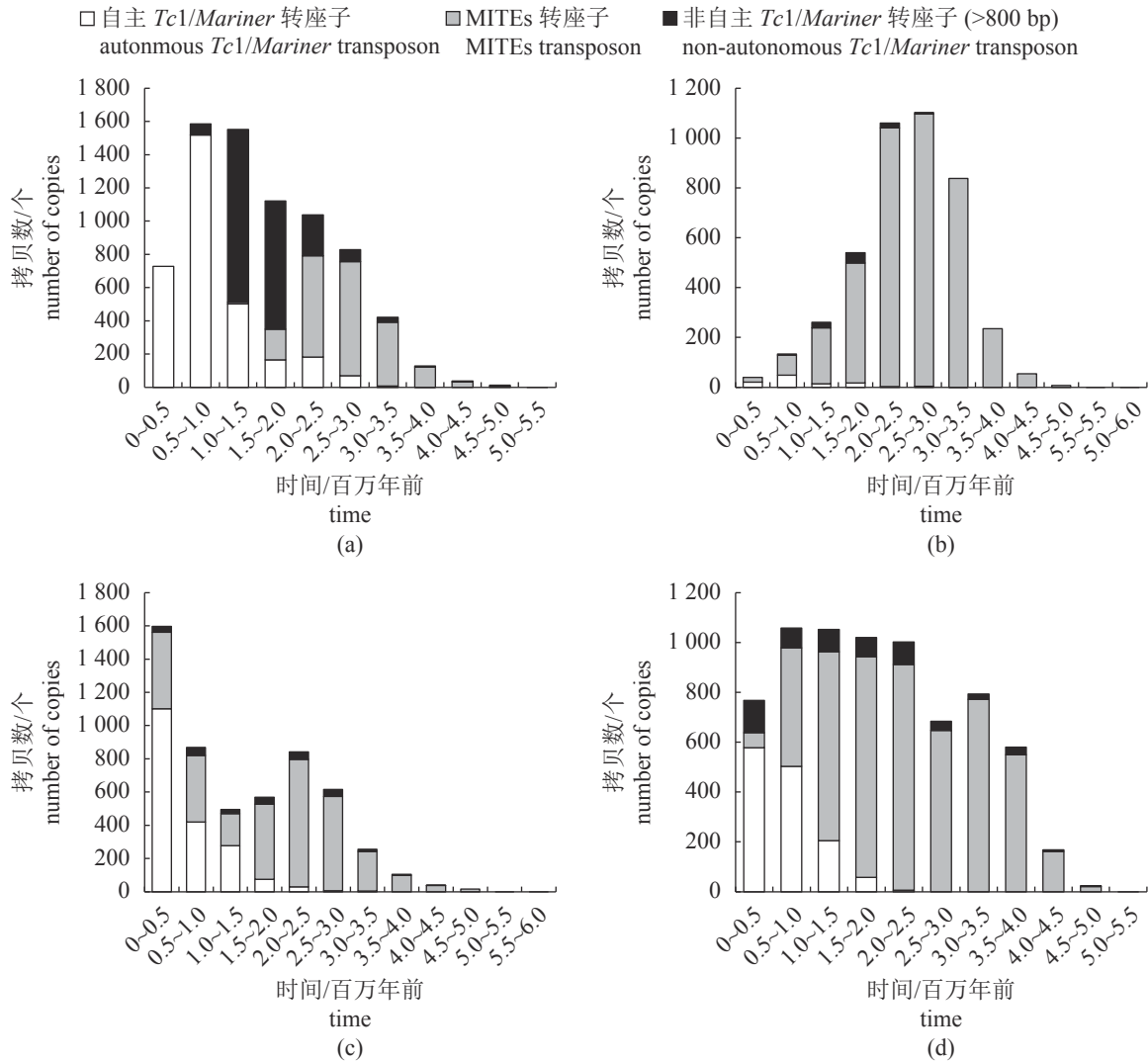


图4 4种鲈基因组中 *Tc1/Mariner* 转座子的插入时间估计

(a) 沟鲈, (b) 黑斑原鲈, (c) 鲈, (d) 南方鲈

Fig. 4 Estimation of insertion time of *Tc1/Mariner* transposons in four catfishes

(a) *I. punctatus*, (b) *G. maculatum*, (c) *S. asotus*, (d) *S. meridionalis*

3.3 *Tc1/Mariner* 转座子的水平转移和分布特征

通过分析发现, 鲈和南方鲈的 *Tc1/Mariner* 转座子系统发育、扩增模式和插入位置等方面的特点相比沟鲈和黑斑原鲈更为类似。这些结果与鲈的垂直演化关系相符合。比较特殊的是沟鲈中的 *Tc1/Mariner-3-Ip(pogo)* 转座子, 实验中在其他3种鲈中并未找到。系统发育分析发现此转座子与红鳍东方鲈、斑马鱼和斑点雀鲈的 *pogo* 转座子聚为一个类群。通过NCBI中的多种数据库比对, 发现 *Tc1/Mariner-3-Ip* 转座子与黄鳝 (*Monopterus albus*) 和眼斑双锯鱼 (*Amphiprion ocellaris*) 基因组序列有极高的相似性(相似度 >80%、覆盖度 >80%、 e 值 < 10^{-5}), 尤其在黄鳝中

<http://www.sxuebao.cn>

覆盖度和相似度都达到90%以上。由此推测, *Tc1/Mariner-3-Ip* 转座子可能是水平转移到沟鲈、黄鳝和眼斑双锯鱼物种中。水平转移也可能是鱼类物种间演化的重要源动力。另外, 通过对插入位置分析发现, *Tc1/Mariner-3-Ip* 转座子家族的全长拷贝主要插入到与血液系统和中枢神经系统等有关的 *VPS13B (COH1)* 基因和与脂质等代谢有关的 *SUGCT* 基因的内含子中^[40-41]。实际上, 转座子在基因组中并不是随机分布, 而是偏好性地插入到某些特定的区域, 这种平衡行为能够促进转座子未来的传播, 同时减轻对宿主细胞功能的有害影响^[42], 例如转座子可以插入到新的调控元件、外显子和内含子中介导基因融合和基因沉默^[43]。很多转座子如 *Tc1* 可以反过来促

表 2 *Tc1/Mariner* 转座子在基因内及侧翼的分布Tab. 2 Distribution of *Tc1/Mariner* transposons in genes and flanks

物种 species	类型 type	5'-侧翼 5'-flank	外显子 exon	内含子 intron	3'-侧翼 3'-flank
沟鲈 <i>I. punctatus</i>	总 <i>Tc1/Mariner</i>	0	31	9 798	0
	自主 <i>Tc1/Mariner</i>	0	9	4 319	0
	MITEs	0	16	2 826	0
黑斑原鲈 <i>G. maculatum</i>	总 <i>Tc1/Mariner</i>	0	4	1 950	0
	自主 <i>Tc1/Mariner</i>	0	3	61	0
	MITEs	0	1	1 850	0
鲈 <i>S. asotus</i>	总 <i>Tc1/Mariner</i>	883	100	2 200	920
	自主 <i>Tc1/Mariner</i>	362	64	714	367
	MITEs	482	32	1 418	526
南方鲈 <i>S. meridionalis</i>	总 <i>Tc1/Mariner</i>	929	87	3 229	1 021
	自主 <i>Tc1/Mariner</i>	241	41	564	250
	MITEs	608	38	2 455	690

使新基因的产生, 从而可以使物种适应新环境^[44-45]。本研究表明4种鲈中的*Tc1/Mariner*转座子都偏向于插入到基因的内含子中。基于本研究采用的侧翼统计范围(<2.5 kb)发现, 沟鲈和黑斑原鲈基因的侧翼未发现*Tc1/Mariner*转座子的插入, 但不排除该转座子可能存在于基因的侧翼位置。另外, 在外显子区域也有所发现, 尤其在鲈和南方鲈中数量颇多。沟鲈、鲈和南方鲈作为底栖生活的鱼类, 更易暴露于各种生物和非生物的胁迫中, 但其均具有个体大、生长快、抗病力强等特点^[14]。据报道, *Tc1/Mariner*转座子似乎已经整合到免疫球蛋白的重复C μ 区^[46]。据此推测, 沟鲈、鲈和南方鲈基因组中*Tc1/Mariner*转座子如此高的含量和分布的偏好性特点可能在其适应逆境中发挥着重要作用。总之, 有关转座子对生物体表型、生理或行为活动等方面的影响还需要后续深入地研究。

4 结论

通过对4种鲈基因组中*Tc1/Mariner*转座子进行鉴定、统计和家族分类, 发现沟鲈、鲈和南方鲈中*Tc1/Mariner*转座子的含量和家族数目相对较高。4种鲈的*Tc1/Mariner*转座子总共可以分为5个不同的亚类群, 其中多数为*Tc1*转座子。插入时间分析表明, 多数*Tc1/Mariner*转座子在0~1百万年前或2~3百万年前有一个突然的扩增, 并且在近期可能存在活性。另外, 4种鲈的*Tc1/Mariner*

转座子更倾向于插入到基因的内含子中。以上研究明确了4种鲈基因组中*Tc1/Mariner*转座子的特征, 为后续鱼类*Tc1/Mariner*转座子的活性和功能等方面的深入研究奠定了基础。

参考文献:

- [1] Robertson H M, Lampe D J. Distribution of transposable elements in arthropods[J]. *Annual Review of Entomology*, 1995, 40: 333-357.
- [2] McClintock B. The origin and behavior of mutable loci in maize[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1950, 36(6): 344-355.
- [3] Wicker T, Sabot F, Hua-Van A, *et al.* A unified classification system for eukaryotic transposable elements[J]. *Nature Reviews Genetics*, 2007, 10(4): 276.
- [4] Feschotte C, Pritham E J. DNA transposons and the evolution of eukaryotic genomes[J]. *Annual Review of Genetics*, 2007, 41: 331-368.
- [5] Eide D, Anderson P. Transposition of *Tc1* in the nematode *Caenorhabditis elegans*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1985, 82(6): 1756-1760.
- [6] Jacobson J W, Medhora M M, Hartl D L. Molecular structure of a somatically unstable transposable element in *Drosophila*[J]. *Proceedings of the National Academy*

- of Sciences of the United States of America, 1986, 83(22): 8684-8688.
- [7] Plasterk R H A, Izsvák Z, Ivics Z. Resident aliens: the *Tc1/Mariner* superfamily of transposable elements[J]. *Trends in Genetics*, 1999, 15(8): 326-332.
- [8] Shao H G, Tu Z J. Expanding the diversity of the *IS630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons[J]. *Genetics*, 2001, 159(3): 1103-1115.
- [9] Gomulski L M, Torti C, Bonizzoni M, *et al.* A new basal subfamily of *mariner* elements in *Ceratitis rosa* and other tephritid flies[J]. *Journal of Molecular Evolution*, 2001, 53(6): 597-606.
- [10] Zhang H H, Shen Y H, Xiong X M, *et al.* Identification and evolutionary history of the DD41D transposons in insects[J]. *Genes & Genomics*, 2016, 38(2): 109-117.
- [11] Liu Y, Yang G J. *Tc1*-like transposable elements in plant genomes[J]. *Mobile DNA*, 2014, 5: 17.
- [12] Yuan Z H, Liu S K, Zhou T, *et al.* Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments[J]. *BMC Genomics*, 2018, 19: 141.
- [13] Peccoud J, Loiseau V, Cordaux R, *et al.* Massive horizontal transfer of transposable elements in insects[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(18): 4721-4726.
- [14] Yuan Z H, Zhou T, Bao L S, *et al.* The annotation of repetitive elements in the genome of channel catfish (*Ictalurus punctatus*)[J]. *PLoS One*, 2018, 13(5): e0197371.
- [15] Huang Y, Niu B F, Gao Y, *et al.* CD-HIT Suite: a web server for clustering and comparing biological sequences[J]. *Bioinformatics*, 2010, 26(5): 680-682.
- [16] Edgar R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput[J]. *Nucleic Acids Research*, 2004, 32(5): 1792-1797.
- [17] Xia X, Xie Z. DAMBE: Software package for data analysis in molecular biology and evolution[J]. *Journal of Heredity*, 2001, 92(4): 371-373.
- [18] Okonechnikov K, Golosova O, Fursov M, *et al.* Unipro UGENE: a unified bioinformatics toolkit[J]. *Bioinformatics*, 2012, 28(8): 1166-1167.
- [19] Tamura K, Stecher G, Peterson D, *et al.* MEGA6: Molecular evolutionary genetics analysis version 6.0[J]. *Molecular Biology and Evolution*, 2013, 31(12): 2725-2729.
- [20] Li W H. *Molecular evolution*[M]. Sunderland: Sinauer Associates Incorporated, 1997.
- [21] Spivakov M, Auer T O, Peravali R, *et al.* Genomic and phenotypic characterization of a wild medaka population: Towards the establishment of an isogenic population genetic resource in fish[J]. *G3: Genes, Genomes, Genetics*, 2014, 4(3): 433-445.
- [22] Nandi S, Peatman E, Xu P, *et al.* Repeat structure of the catfish genome: a genomic and transcriptomic assessment of *Tc1*-like transposon elements in channel catfish (*Ictalurus punctatus*)[J]. *Genetica*, 2007, 131(1): 81-90.
- [23] Liu Z, Li P, Kucuktas H, *et al.* Characterization of nonautonomous *Tc1*-like transposable elements of channel catfish (*Ictalurus punctatus*)[J]. *Fish Physiology and Biochemistry*, 1999, 21(1): 65-72.
- [24] Vos J C, Plasterk R H S. *Tc1* transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain[J]. *The EMBO Journal*, 1994, 13(24): 6125-6132.
- [25] Gao B, Chen W, Shen D, *et al.* Characterization of autonomous families of *Tc1/Mariner* transposons in neoteleost genomes[J]. *Marine Genomics*, 2017, 34: 67-77.
- [26] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences[J]. *Journal of Molecular Evolution*, 1980, 16(2): 111-120.
- [27] Santiago N, Herráiz C, Goñi J R, *et al.* Genome-wide analysis of the *Emigrant* family of MITES of *Arabidopsis thaliana*[J]. *Molecular Biology and Evolution*, 2002, 19(12): 2285-2293.
- [28] Aziz R K, Breitbart M, Edwards R A. Transposases are the most abundant, most ubiquitous genes in nature[J]. *Nucleic Acids Research*, 2010, 38(13): 4207-4217.
- [29] Lohe A R, Moriyama E N, Lidholm D A, *et al.* Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements[J]. *Molecular Biology and Evolution*, 1995, 12(1): 62-72.
- [30] Chalopin D, Naville M, Plard F, *et al.* Comparative analysis of transposable elements highlights mobilome

- diversity and evolution in vertebrates[J]. *Genome Biology and Evolution*, 2015, 7(2): 567-580.
- [31] 张化浩. 家蚕中 *Tc1/Mariner* 转座子的全基因组鉴定及进化分析[D]. 重庆: 西南大学, 2012.
- Zhang H H. Genome-wide identification and evolution of *Tc1/Mariner* in the silkworm (*Bombyx mori*) genome[D]. Chongqing: Southwest University, 2012(in Chinese).
- [32] Bouallègue M, Filée J, Kharrat I, *et al.* Diversity and evolution of *mariner*-like elements in aphid genomes[J]. *BMC Genomics*, 2017, 18(1): 494.
- [33] Puzakova L V, Puzakov M V. The *Tc1/mariner* DNA transposons in the genome of mollusk *Littorina saxatilis*[J]. *Russian Journal of Genetics*, 2017, 53(12): 1358-1365.
- [34] Horváth V, Merenciano M, González J. Revisiting the relationship between transposable elements and the eukaryotic stress response[J]. *Trends in Genetics*, 2017, 33(11): 832-841.
- [35] Clark K J, Carlson D F, Leaver M J, *et al.* *Passport*, a native *Tc1* transposon from flatfish, is functionally active in vertebrate cells[J]. *Nucleic Acids Research*, 2009, 37(4): 1239-1247.
- [36] Pujolar J M, Astolfi L, Boscarì E, *et al.* *Tana1*, a new putatively active *Tc1*-like transposable element in the genome of sturgeons[J]. *Molecular Phylogenetics and Evolution*, 2013, 66(1): 223-232.
- [37] Guo X M, Zhang Q Q, Sun Y W, *et al.* *Tc1*-like transposase *Thm3* of silver carp (*Hypophthalmichthys molitrix*) Can mediate gene transposition in the genome of blunt snout bream (*Megalobrama amblycephala*)[J]. *G3: Genes, Genomes, Genetics*, 2015, 5(12): 2601-2610.
- [38] Bureau T E, Wessler S R. Tourist: a large family of small inverted repeat elements frequently associated with maize genes[J]. *The Plant Cell*, 1992, 4(10): 1283-1294.
- [39] Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution[J]. *Trends in Ecology and Evolution*, 2010, 25(9): 537-546.
- [40] Balikova I, Lehesjoki A E, de Ravel T J, *et al.* Deletions in the *VPS13B (COH1)* gene as a cause of Cohen syndrome[J]. *Human Mutation*, 2009, 30(9): E845-E854.
- [41] La Serna-Infantes J, Pastor M C, Trubnykova M, *et al.* Novel contiguous gene deletion in peruvian girl with trichothiodystrophy type 4 and glutaric aciduria type 3[J]. *European Journal of Medical Genetics*, 2018, 61(7): 388-392.
- [42] Bourque G, Burns K H, Gehring M, *et al.* Ten things you should know about transposable elements[J]. *Genome Biology*, 2018, 19: 199.
- [43] Buckley R M, Adelson D L. Mammalian genome evolution as a result of epigenetic regulation of transposable elements[J]. *Biomolecular Concepts*, 2014, 5(3): 183-194.
- [44] González J, Lenkov K, Lipatov M, *et al.* High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*[J]. *PLoS Biology*, 2008, 6(10): e251.
- [45] González J, Macpherson J M, Petrov D A. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*[J]. *Molecular Biology and Evolution*, 2009, 26(9): 1949-1961.
- [46] Ventura-Holman T, Lobb C J. Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the IgH locus represents a composite of two gene clusters[J]. *Molecular Immunology*, 2002, 38(7): 557-564.

Identification and evolutionary analysis of *Tc1/Mariner* transposons in four catfish genomes

ZHAO Lianpeng, SHAO Feng, HU Jingwen, ZHANG Yaoguang, PENG Zuogang*

(Key Laboratory of Freshwater Fish Reproduction and Development, Ministry of Education,
School of Life Sciences, Southwest University, Chongqing 400715, China)

Abstract: To investigate the characteristics of *Tc1/Mariner* transposons in four catfish genomes (*Ictalurus punctatus*, *Glyptosternum maculatum*, *Silurus asotus* and *S. meridionalis*) and their role in the evolution, we used *de novo* and homology methods to predict *Tc1/Mariner* transposons and explored the role in the evolution of those four catfishes. The results showed that the contents of *Tc1/Mariner* transposons in the four catfishes were 9.46%, 2.70%, 7.83% and 8.54%, respectively. All the *Tc1/Mariner* transposons in the four catfishes were classified into 38, 20, 43 and 55 families, respectively. Based on the structural and phylogenetic analysis of the catalytic region of the transposase, the *Tc1/Mariner* transposons in the four catfishes can be classified into five different subgroups: *Tc1* (DD34E), DD35E, DD36E, DD37E and *pogo* (DD35D). Phylogenetic analysis indicated that the transposons of most of the branches for *S. asotus* and *S. meridionalis* were clustered into sister groups. The results indicated that the insertion time of the majority *Tc1/Mariner* transposons mainly ranges between 0-5 Mya (million years ago). Many *Tc1/Mariner* transposons have a burst expansion between 0-1 Mya or 2-3 Mya. In addition, most of the *Tc1/Mariner* transposons in the four catfishes were inserted into the intron of the gene. In summary, this study provides new insights into the evolution of *Tc1/Mariner* transposons in the fish genomes.

Key words: catfish; genome; transposon; *Tc1/Mariner*; evolutionary analysis

Corresponding author: PENG Zuogang. E-mail: pzg@swu.edu.cn

Funding projects: National Natural Science Foundation of China (31572254)