

柞蚕微孢子虫基因组 LTR 反转座子鉴定与进化分析^{*}

周俊, 石鹏, 张爽, 许金山, 张小燕

(重庆师范大学 重庆市媒介昆虫重点实验室 重庆市动物生物学重点实验室, 重庆 401331)

摘要:【目的】鉴定柞蚕微孢子虫(*Nosema pernyi*)基因组中 LTR 反转座子的种类、数量和遗传关系。【方法】从 GenBank 数据库下载柞蚕微孢子虫基因组序列,采用 MGEScan 软件筛选潜在重复元件,通过 NCBI-BLAST 与不同物种的逆转酶进行比对,含有逆转录酶结构域的序列认定为 LTR 反转座子,并探讨它们的结构特征、所属类型和基因组分布特点。【结果】从柞蚕微孢子虫基因组中共发现 6 个 LTR 反转座子家族即 Nar1, Nar2, Nar3, Nar4, Nar5 和 Nar6, 长度在 3~4 kb 之间;大部分家族有相似的 LTR 末端核苷酸,但靶位点重复和拷贝数存在差异;逆转录酶结构域系统进化树显示 6 个 LTR 反转座子家族均属于 Ty3/Gypsy 类型,且与家蚕微孢子虫(*Nosema bombycis*)LTR 反转座子形成姊妹分支;Nar2 和 Nar4 家族在基因组上成串排列,具有相似的分布特征。【结论】柞蚕微孢子虫基因组中存在 Ty3/Gypsy 类型的 LTR 反转座元件,进化地位与家蚕微孢子虫最为接近;LTR 反转座子家族重复元件在基因组上成串排列,可能是柞蚕微孢子虫基因组扩张的潜在因素之一。

关键词: LTR 反转座子;柞蚕微孢子虫;基因组;系统进化

中图分类号: Q175

文献标志码: A

文章编号: 1672-6693(2019)03-0029-07

转座因子是一段可以从染色体基因组某一位点转移到另一位点并影响转座位点周围相关基因功能的 DNA 序列,研究发现至少有 40% 以上的人类基因组序列为转座因子,而在一些植物基因组中,这一比例达到 70% 以上^[1-2]。根据转座机制的不同和是否形成 RNA 转座中间体,可将转座子主要分为两大类:一类为 DNA 转座子,主要通过剪切-粘贴模式完成自身在基因组上的转座;第二类为反转座子,主要以复制-粘贴模式通过转录和反转录过程形成新的转座子拷贝,并插入到基因组新的转座位点从而完成转座过程。反转座子可以划分为 LTR 反转座子(Long terminal direct repeat retrotransposon)和 Non-LTR 反转座子。LTR 反转座子主要包含 5 个超家族:Ty1/Copia (Pseudoviridae), Ty3/Gypsy (Metaviridae), Bel/Pao, Retroviral 和 ERV (Endogenous retrovirus)^[3]。一般而言,转座因子选择基因组的某一插入位点进行转座时,会导致插入该点处的一段短核苷酸片段发生自我复制;就 LTR 反转座子而言,该片段长度通常在 4~6 bp 之间,被称之为靶位点重复(Target site duplication),它的存在已成为转座因子的特性之一。早期研究认为转座因子是一种自私 DNA 序列(Selfish DNA)或者垃圾 DNA 序列(Junk DNA),但越来越多的研究表明,转座因子对基因组和基因进化起到积极作用,在许多动物和植物中,转座子不仅能通过特殊的途径重组基因组,而且能转座入完整基因的内部,进而影响基因的结构及功能^[4-7]。

柞蚕微孢子虫(*Nosema pernyi*)是引起柞蚕微孢子病的主要病原体,也是柞蚕生产中重要的检疫对象^[8],主要通过经口传染和胚种传染两种传播途径感染柞蚕^[9],极易造成柞蚕丝茧减产,危害中国的柞蚕生产。关于微孢子虫进化地位研究一向是人们关注的焦点。早先研究者根据微孢子 SSUrRNA 系统进化分析并结合微孢子缺乏线粒体和过氧化物酶特点,认为微孢子属于原生生物^[10]。而后更多分子生物学证据,比如发现 5S rRNA 为典型真核生物型, β -微管蛋白和 RNA 聚合酶 II 大亚基与真菌最为接近,发现线粒体型 HSP70 编码基因等,使人

* 收稿日期:2019-03-10 修回日期:2019-04-12 网络出版时间:2019-05-09 19:30

资助项目:重庆市基础与前沿研究计划项目(No. cstc2015jcyjA80020)

第一作者简介:周俊,男,研究方向为生物化学与分子生物学, E-mail: sjzhou2010@163.com;通信作者:张小燕,女,副教授, E-mail: xy Zhang18@163.com

网络出版地址: <http://kns.cnki.net/kcms/detail/50.1165.N.20190509.1930.044.html>

们提出了微孢子是真菌或真菌的类似物中分化而来,微孢子虫更接近真菌的新观点^[11-13]。

现阶段在基因组水平上研究转座子而涉及的简单真核生物中,多数与真菌类有关,且其中的 LTR 反转座子主要属于 Ty3/Gypsy 类^[14]。在新型隐球菌(*Cryptococcus neoformans*)基因组中,有 15 个 LTR 反转座子家族和几个 Non-LTR 反转座子家族被发现^[15]。在白色念珠菌(*Candida albicans*)基因组中则至少有 34 个反转座子家族被发现^[16],而且高温环境和抑菌剂咪康唑能分别诱导其中的 Tac4 家族与 Zorro2 家族转录激活,从而增强了该菌种的适应性^[17]。此外,在家蚕微孢子虫(*Nosema bombycis*)基因组中相继发现 14 个结构完整的 LTR 反转座子家族以及其他类型的转座子^[18-19],而后续研究则揭示转座子家族的扩增是家蚕微孢子虫基因组扩张的原因之一^[20]。目前关于柞蚕微孢子虫反转座子的报道很少,而且只报道了柞蚕微孢子虫基因组中存在 1 个 LINE 反转座子家族 R4Na^[21],但有关 LTR 反转座子的研究罕有报道。为此,本研究利用生物信息学技术,发掘柞蚕微孢子虫基因组中存在的 LTR 反转座子,为进一步理解 LTR 反转座子在柞蚕微孢子虫基因组进化中的作用提供基础依据。

1 材料与方法

1.1 LTR 反转座元件的鉴定

本研究所使用的柞蚕微孢子虫全基因组和家蚕微孢子虫基因组下载自 Genbank 公共数据库,序列收录号为:ACJZ01000001~ACJZ01003558。首先,采用 MGEScan^[22] 搜索柞蚕微孢子虫全基因组序列,鉴定潜在的 LTR 反转座元件。再从 Pfam 数据库下载多物种的反转座酶(Reverse transcriptase, RT)序列,建库后采用 NCBI-BLAST 比对候选 LTR 反转座元件,确定是否含有 RT 序列。本研究将完整的 LTR 反转座子定义为序列内含有 RT 基因且两端拥有长末端重复(Long terminal repeat, LTR)。家族分类参照前人采用的标准,即不同 LTR 反转座子成员内 RT 结构域的氨基酸相似性达到 90% 以上就视为同一家族^[23]。采用 Repeatmasker 重复序列查找软件^[24] 分析特定转座子家族在基因组上的拷贝个数;拷贝片段长度在 100 bp 以上的,相似性达到 90% 以上,定义为 1 个拷贝。多重序列的比对使用 ClustalX2 软件^[24],可视化结果由 BoxShade 3.21 软件生成^[25]。

1.2 系统进化树分析

从 GenBank 公共数据库下载现有 Ty1/Copia, Ty3/Gypsy, Bel/Pao 和 Retroviral 的 LTR 反转座子超家族中反转录酶结构域序列,结合本研究中已鉴定到的柞蚕微孢子虫 LTR 反转座子中的 RT 结构域序列,采用 MEGA 6.0 软件^[26] 建立 Neighbor-joining 系统进化树,选用 No. of differences 模型且 bootstrap 值设为 100。

1.3 LTR 反转座元件分布特征分析

以拷贝数较少和拷贝数较多的 LTR 反转座子家族为例,将两者及它们的拷贝子通过 NCBI-BLAST 与柞蚕微孢子虫基因组进行比对,靶定两个 LTR 反转座子家族及它们的拷贝子在基因组上的位置。靶定特定转座子及拷贝子位置后,查询基因组图谱上下游临近基因,并绘制特定 LTR 反转座子及基因分布图。

2 结果与分析

2.1 LTR 反转座元件鉴定和结构分析

LTR 反转座子主要编码 Pol 多聚蛋白来完成转座功能,该蛋白包含了多个催化结构域,其中以保守性强的 RT 为最重要特征。根据这一特点并结合反转座子结构的特异性,本研究在柞蚕微孢子虫基因组中鉴定到 6 个 LTR 反转座子家族,分别命名为 Nar1, Nar2, Nar3, Nar4, Nar5 和 Nar6。其中, Nar1, Nar2, Nar3, Nar4 和 Nar5 家族含有完整的转座子家族所具有的特征性序列即反向末端重复(Inverted terminal repeat)和靶位点重复序列(表 1)。研究结果还显示:柞蚕微孢子虫 LTR 反转座子的长度在 3~4 kb 之间;大部分家族的 LTR 末端核苷酸具有一致性,为 TGT/ACA 或 TG/CA;但家族间的靶位点重复序列在核苷酸数量和种类上则具有差异,长度为 5~6 个核苷酸(表 1)。此外,通过 RepeatMasker 软件对 LTR 转座子在柞蚕微孢子虫基因组上的拷贝个数进行了初步的统计,发现不同家族间拷贝子数量相差较大,其中拷贝子数量最多的为 Nar3 和 Nar4 家族,均为 16 个;而 Nar2 家族拷贝子数量最少,只有 3 个(表 1)。

表 1 柞蚕微孢子虫基因组 LTR 反转座子不同家族特征详细统计表

Tab. 1 Characteristics of long-terminal repeat retrotransposons from *N. pernyi* genome

家族	LTR 长度/bp	插入片段长度/bp	LTR 相似性/%	靶位点重复序列	LTR 末端核苷酸	拷贝数/个
Nar1	177/177	4 016	100	CTTATT/CTTATT	TGT/ACA	5
Nar2	281/280	4 352	98	AAAAG/AAAAG	TG/CA	3
Nar3	213/202	3 999	100	CCAAT/CCAAT	TG/CA	16
Nar4	177/132	3 036	94	TTTGT/GTTGT	TG/CA	16
Nar5	131/131	3 779	96	ATCTT/ATCTT	TGT/ACA	11
Nar6	451/451	2 705	99	—	—	4

在本研究鉴定到的 6 个 LTR 反转座子家族中,LTR 序列和功能结构域表现出特征差异(图 1)。表 1 显示:LTR 长度范围为 131~451 bp,其中 LTR 长度最长和最短的 LTR 反转座子家族分别为 Nar6 和 Nar5。图 1 则显示,Nar1 和 Nar6 家族含有 RT 结构域和核糖核酸内切酶(RH)结构域共 2 个结构域,而 Nar4 含有 Gag 结构域、蛋白酶(Pro)结构域、RT 结构域和整合酶(INT)结构域共 4 个结构域。尽管柞蚕微孢子虫 LTR 反转座子 pol 多聚蛋白包含的功能结构域具有差异性,但已鉴定的 6 个 LTR 反转座子家族中均具有保守性较强的 RT 结构域,且所有 LTR 反转座子家族的 RT 结构域含有 YXDD 保守基序(图 2a);RH 结构域中则拥有 DAS 保守基序(图 2b);Pro 结构域具有共有的 D(T/S)G 基序(图 2c)。不同 LTR 反转座子家族在基因组上拷贝数差异较大,某些家族 LTR 的相似性差异较大,如 Nar4 家族的为 94%(表 1)。因此推测:LTR 反转座子在较为久远的年代插入到柞蚕微孢子虫基因组中,碱基经过长时间的进化突变,导致高变异度序列不断积累,并伴随有反转座子部分结构域丢失。

2.2 系统进化分析

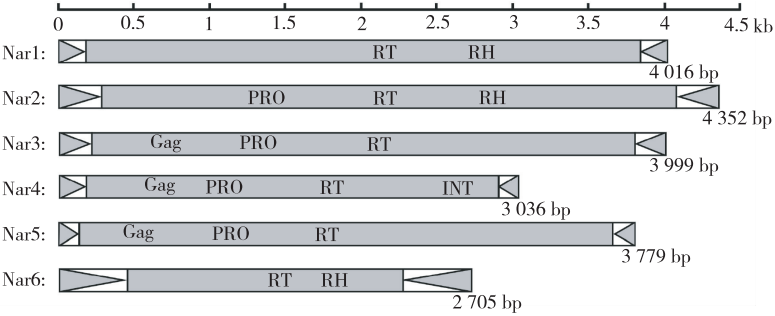
利用柞蚕微孢子虫 LTR 反转座元件中 RT 结构域的保守性,将该结构域与其他物种的 RT 序列构建系统进化树,其中所涉及的各物种的 RT 序列名称、所属物种拉丁学名及 Genbank 登录号如图 3 所示。图 3 还显示:本研究所有物种基因组中 LTR 转座元件可以划分为 Ty3/Gypsy, Retrovirus, Ty1/Copia 和 Bel/Pao 等 4 个类群,与前人报道一致^[27];柞蚕微孢子虫中新发现的 6 个 LTR 反转座子家族均属于 Ty3/Gypsy 类型,且与近源物种家蚕微孢子虫基因组中的 Ty3 类型反转座子聚类在一起形成姊妹分支,暗示柞蚕微孢子虫 LTR 反转座子起源与家蚕微孢子虫 LTR 反转座子相似,属于真菌性起源^[13]。

2.3 LTR 反转座子在基因组中的分布特征

选取拷贝数较少的 Nar2 家族和拷贝数较多的 Nar4 家族为代表,对它们的反转座子及拷贝子进行全基因组定位,考察柞蚕微孢子虫 LTR 反转座子在整个基因组的分布特征。结果显示,Nar2 家族在 Scaffold NA_040 上有 2 个拷贝,在基因组上呈现出紧密排列的特征,并且与其他 Nar 家族成簇分布(图 4a)。Nar4 家族反转座元件在 Scaffold NA_051 上紧密分布 2 个拷贝,且与其他 4 个 Nar 家族成簇排列(图 4b)。推测是由于反转座子在基因组转座过程中受到进化选择,趋向于集中在一起,以防止基因组产生较大变异^[28-29]。

3 讨论

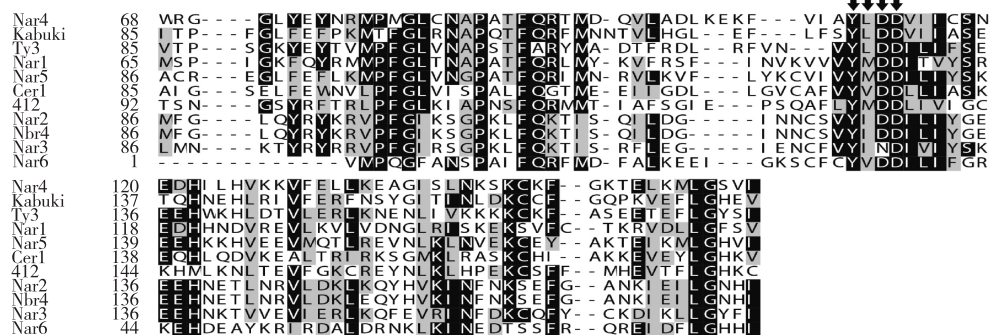
柞蚕微孢子为寄生于柞蚕体内的一类原始真核生物,在微孢子进化地位的研究上具有重要意义。本研究鉴定到 Nar1,Nar2,Nar3,Nar4,Nar5 和 Nar6 共 6 个柞蚕 LTR 反转座子家族,长度在 3~4 kb 之间,证实柞蚕微



注:带有三角形的框表示 LTR,阴影框表示开放阅读框(ORF),并用英文字母标出反转座子各个功能结构域在 ORF 内潜在的位置

图 1 柞蚕微孢子虫不同家族 LTR 反转座子结构

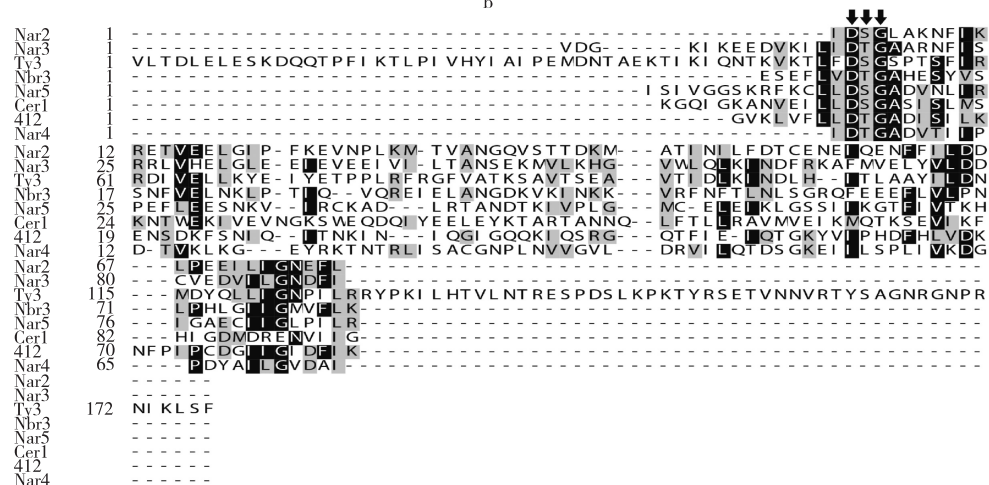
Fig. 1 The presentation of structure for several novel *N. pernyi* LTR retrotransposons

[illegible]

8



b



c

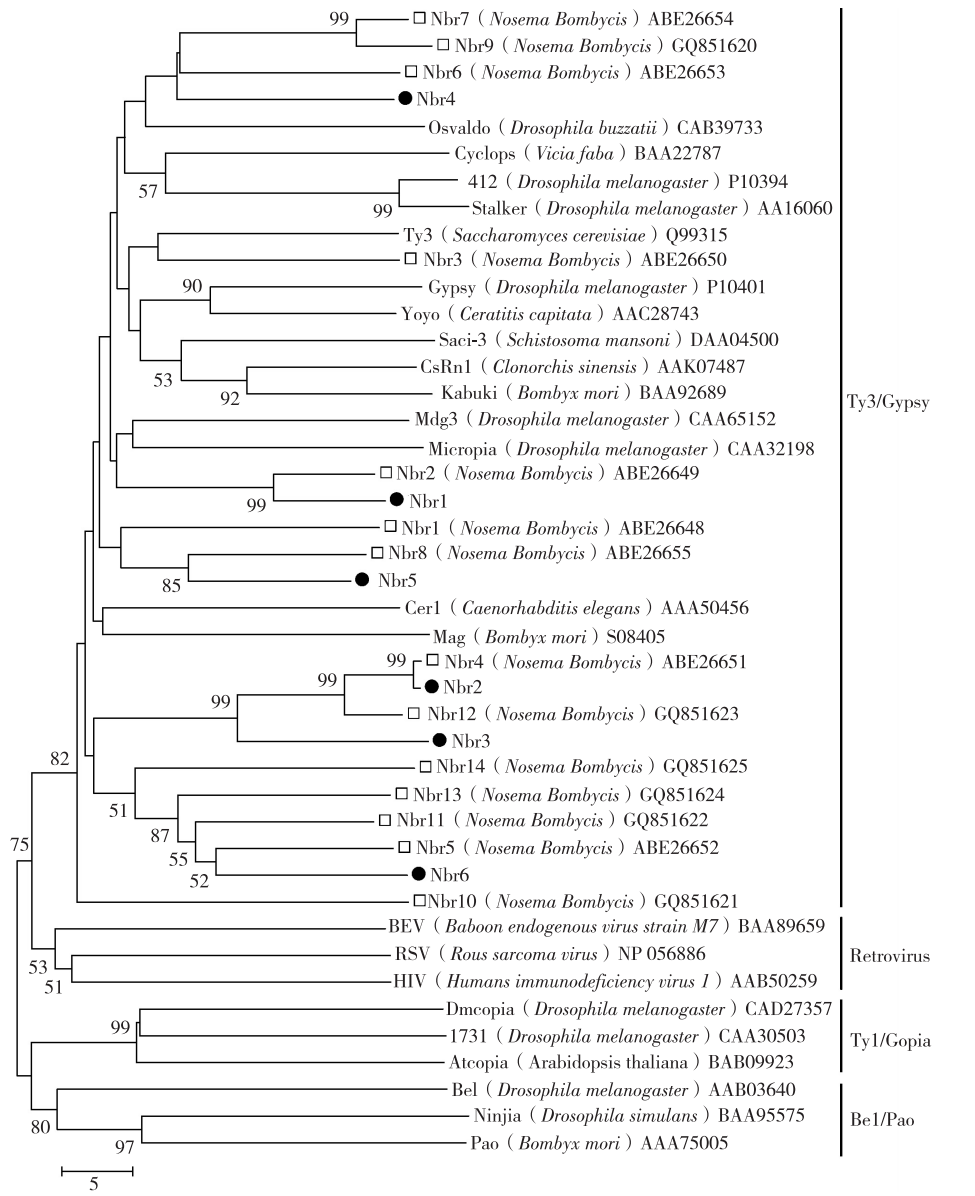
gans, *Drosophila melanogaster* 等物种中 LTR 反转座子序列

Fig. 2 The presentation of multiple alignments for *N. pernyi* LTR retrotransposons

反转座子虽然广泛分布于微孢子虫基因组中,但被验证的具有活性的反转座子报道不多。在家蚕微孢子虫基因组中鉴定到了 7 个有活性的 LTR 反转座子家族,分别为 Nbr5, Nbr9, Nbr10, Nbr11, Nbr12, Nbr13 和

Nbr14^[30],这些家族中反转座子聚合酶基因存在结构域缺失或者插入冗余序列的情况。对于反转座子家族中结构域缺失但仍保留活性的现象,有学者提出,LTR 反转座子家族中各成员的结构域在功能上具有互补性^[31]。在柞蚕微孢子虫 LTR 反转座子中,不同家族间呈现结构多样性,某些家族缺失部分功能结构域,但 RT 结构域、RH 结构域和 Pro 结构仍存在保守基序。本研究推测柞蚕微孢子虫部分 LTR 反转座子家族 ORF 结构的缺失并不会导致活性完全丧失。系统进化树则显示柞蚕微孢子虫中 LTR 反转座子 Nar6 家族与家蚕微孢子虫有转录活性的 Nbr5, Nbr11, Nbr12, Nbr13 和 Nbr14 家族聚为一支,这暗示 Nar6 可能具有转录活性,但柞蚕微孢子虫 LTR 反转座子家族是否还具有转座活性有待进一步研究。

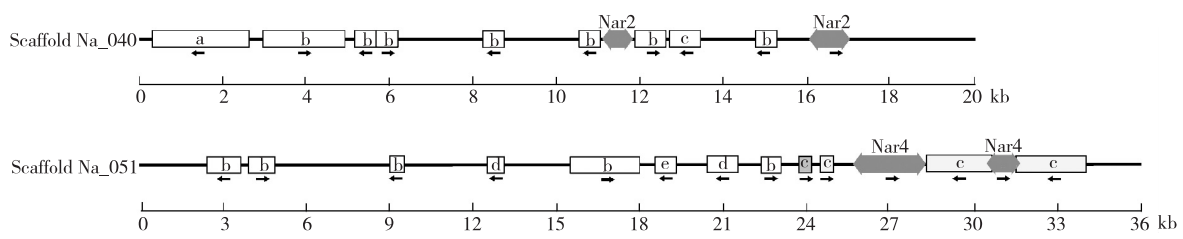
此外, Nar2 和 Nar4 家族在柞蚕基因组中成串分布,并且与其他转座子成簇聚集,与前人报道的微孢子虫反转座子在基因组内具有特定排布方式相一致。如家蚕微孢子虫 LTR 反转座子通常与其他转座元件聚集并插入到两个共线性区域^[19,21]。柞蚕微孢子虫 LTR 反转座子相对集中的分布方式,减轻了对整个基因组产生突变的危害^[28-29]。



注:圆圈表示柞蚕微孢子虫 LTR 反转座子家族,空心方框表示家蚕微孢子虫已确定的 LTR 反转座子家族,括号内为物种拉丁学名,括号后方为 GenBank 登录号

图 3 柞蚕微孢子虫和其他物种基于 LTR 反转座元件中 RT 结构域序列的系统进化树

Fig. 3 Phylogenetic tree of the Nar retrotransposons and homologues in other species based on their reverse transcriptase sequences



注:方框中的字母 a~e 分别表示几丁质合成酶、假设蛋白、其他 Nar 家族、无特征蛋白和转座酶

图 4 LTR 反转座子在柞蚕微孢子虫基因组中的分布

Fig. 4 Distribution of LTR element in *N. pernyi* genomic

参考文献:

- [1] SMIT A F. Interspersed repeats and other mementos of transposable elements in mammalian genomes[J]. *Current Opinion in Genetics & Development*, 1999, 9(6): 657-663.
- [2] SANMIGUEL P, TIKHONOV A, JIN Y K, et al. Nested retrotransposons in the intergenic regions of the maize genome[J]. *Science*, 1996, 274(5288): 765-768.
- [3] POULTER R T M, GOODWIN T J D, DIRS-1 and the other tyrosine recombinase retrotransposons[J]. *Cytogenetic and Genome Research*, 2005, 110(1/2/3/4): 575-588.
- [4] KUMAR A, BENNETZEN J L. Plant retrotransposons[J]. *Annual Review of Genetics*, 1999, 33: 479-532.
- [5] KIDWELL M G, LISCH D R. Transposable elements and host genome evolution[J]. *Trends in Ecology and Evolution*, 2000, 15(3): 95-99.
- [6] GANKO E W. Evolutionary history of *Cer* elements and their impact on the *C. elegans* genome[J]. *Genome Research*, 2001, 11(12): 2066-2074.
- [7] GAO L, Mc CARTHY E M, GANKO E W, et al. Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences[J]. *BMC Genomics*, 2004, 5(1): 18.
- [8] 秦利. 中国柞蚕学[M]. 北京: 中国科学文化出版社, 2003: 347-352.
QIN L. Chinese tussah[M]. Beijing: China Science and Culture Press, 2003: 347-352.
- [9] 王勇. 柞蚕微孢子虫 EST 分析、孢壁蛋白 *NpSWP1* 基因克隆及柞蚕免疫应答研究[D]. 沈阳: 沈阳农业大学, 2014.
WANG Y. EST analysis of *Nosema pernyi*, *NpSWP1* gene cloning and immune response in *Antheraea pernyi* [D]. Shenyang: Shenyang Agricultural University, 2014.
- [10] VOSSBRINCK C R, MADDOX J V, FRIEDMAN S, et al. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes[J]. *Nature*, 1987, 326(6111): 411-414.
- [11] KEELING P J, LUKER M A, PALMER J D. Evidence from Beta-tubulin phylogeny that microsporidia evolved from within the fungi[J]. *Molecular Biology and Evolution*, 2000, 17(1): 23-31.
- [12] HIRT R P, LOGSDON J M Jr, HEALY B, et al. Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(2): 580-585.
- [13] HIRT R P, HEALY B, VOSSBRINCK C R, et al. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria[J]. *Current Biology*, 1997, 7(12): 995-998.
- [14] 周泽扬. 家蚕微孢子虫基因组生物学[M]. 北京: 科学出版社, 2014.
ZHOU Z Y. Genome biology of *Nosema bombycis* [M]. Beijing: Science Press, 2014.
- [15] GOODWIN T J D, POULTER R T M. The diversity of retrotransposons in the yeast *Cryptococcus neoformans* [J]. *Yeast*, 2001, 18(9): 865-880.
- [16] GOODWIN T J D. Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans* [J]. *Genome Research*, 2000, 10(2): 174-191.
- [17] 苗琦. Tca4 和 Zorro2 的转座激活与白念珠菌适应性形成的相关性研究[D]. 上海: 第二军医大学, 2013.
MIAO Q. Study between the transposition of Tca4 and Zorro2 with the adaptability formation of *Candida albicans* [D]. Shanghai: Second Military Medical University, 2013.
- [18] XU J S, PAN G Q, FANG L, et al. The varying microsporidian genome: existence of long-terminal repeat retrotransposon in domesticated silkworm parasite *Nosema bombycis* [J]. *International Journal for Parasitology*, 2006, 36(9): 1049-1056.
- [19] XIANG H, PAN G Q, ZHANG R Z, et al. Natural selection maintains the transcribed LTR retrotransposons in *Nosema bombycis* [J]. *Journal of Genetics and Genomics*, 2010, 37(5): 305-314.
- [20] PAN G Q, XU J S, LI T, et al. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation[J]. *Bmc Genomics*, 2013, 14(1): 186-186.
- [21] 刘兰兰, 吴春漫, 张小燕, 等. 家蚕微孢子虫和柞蚕微孢子虫 LINE 类反转录转座子 R4 的鉴定及系统进化分析[J]. *蚕业科学*, 2014, 40(4): 681-687.
LIU L L, WU C M, ZHANG X Y, et al. Identification and phylogenetic analysis of LINE retrotransposons R4 in *Nosema bombycis* and *Nosema pernyi* [J]. *Sericultural Science*, 2014, 40(4): 681-687.
- [22] LEE H, LEE M, ISMAIL W M, et al. MGEScan: a galaxy-based system for identifying retrotransposons in genomes [J]. *Bioinformatics*, 2016, 32(16): 2502-2504.
- [23] BOWEN N J, Mc DONALD J F. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside[J]. *Genome research*, 2001, 11(9): 1527-1540.
- [24] CHEN N. Using repeat masker to identify repetitive elements in genomic sequences [J]. *Current Protocols in*

- Bioinformatics, 2004, 5(1): 4-10.
- [25] BODZIOCH M, LAPICKA-BODZIOCH K, ZAPALA B, et al. Evidence for potential functionality of nuclearly-encoded humanin isoforms[J]. Genomics, 2009, 94(4): 247-256.
- [26] TAMURA K, STECHER G, PETERSON D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0[J]. Molecular Biology and Evolution, 2013, 30(12): 2725-2729.
- [27] WICKER T, SABOT F, HUA-VAN A, et al. A unified classification system for eukaryotic transposable elements [J]. Nature Reviews Genetics, 2007, 8(12): 973-982.
- [28] PEREIRA V. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome [J]. Genome Biology, 2004, 5(10): R79.
- [29] BARTOLOME C, MASIDE X, CHARLESWORTH B. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster* [J]. Molecular Biology and Evolution, 2002, 19(6): 926-937.
- [30] 张瑞芝. 家蚕微孢子虫 LTR 反转座子的活性研究[D]. 重庆: 西南大学, 2008.
- ZHANG R Z. Characterization of active LTR retrotransposons in *Nosema bombycis* [D]. Chongqing: Southwest University, 2008.
- [31] SABOT F, SCHULMAN A H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome [J]. Heredity, 2006, 97(6): 381.

Animal Sciences

Identification and Phylogenetic Analysis of LTR Retrotransposons in *Nosema pernyi*

ZHOU Jun, SHI Peng, ZHANG Shuang, XU Jinshan, ZHANG Xiaoyan

(Chongqing key Laboratory of Vector Insects, Chongqing Key Laboratory of Animal Biology,
Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] To identify the species, quantity and genetic relationship of the LTR inverted locus in the genome of *Nosema pernyi*. [Methods] The genomic sequence of *N. pernyi* was downloaded from the GenBank database, and screened to identify the potential LTR retrotransposons based on MGEScan software. Sequence alignment between potential LTR retrotransposons and reverse transcriptase were performed through NCBI-blast software. Those elements containing the reverse transcriptase domain were considered as the LTR retrotransposons. The structural character, classification type and genomic distribution of these LTR retrotransposons were further explored. [Findings] A total of 6 LTR retrotransposons families were found in the *N. pernyi* genome, ranging in length from 3kb to 4kb. Most families have similar LTR terminal nucleotides, but show the sequence difference about target site duplication and copy number. The phylogenetic tree of the RT domain shows that all the Nars belongs to the Ty3/Gypsy type and forms a sister branch with the *Nosema bombycis* LTR retrotransposons. The copies of Nar4 family and Nar2 family were arranged in tandem on the genome with similar distribution characteristics. [Conclusions] The *N. pernyi* genome has several complete Gypsy-type LTR retransposon, which show the closely evolutionary relationship to the elements from the *N. bombycis*. The copies of Nars were arranged in a cluster on the genome and may be one of the potential factors for the expansion of the *N. pernyi* genome.

Keywords: LTR retrotransposons; *Nosema pernyi*; genome; phylogenetic

(责任编辑 方 兴)