

美洲大蠊基因组重复序列分析

牟必琴¹, 严超超¹, 李午佼¹, 李静¹, 沈咏梅², 岳碧松^{1*}

(1. 四川大学生命科学学院, 生物资源与生态环境教育部重点实验室, 成都 610064;

2. 药用美洲大蠊四川省重点实验室, 成都 610081)

摘要: 重复序列是真核生物基因组的重要组成部分。一些重复序列, 如自主型的逆转录转座子 LINE, 在昆虫的系统进化和遗传多样性研究方面得到了广泛的应用。*de novo* 从头预测和基于同源比对预测相结合的方法被用来搜索美洲大蠊 *Periplaneta americana* 基因组, 共鉴定出大约占全基因组 62% 的重复序列。研究发现, 散在重复序列中, DNA 转座子占美洲大蠊基因组的 16.18%; 逆转座元件中 LINE 最多, 占基因组的 13.64%, SINE 和 LTR 逆转录转座子分别占基因组的 3.52% 和 1.32%。LINEs 中的 BovB 亚家族在所有转座子亚家族中比例最高(约 6.73%)。美洲大蠊与德国小蠊 *Blattella germanica* 相比, 除 LTR 逆转座子外, 其他类型的转座子占基因组的比例均高于德国小蠊。通过分析逆转录转座子反转录酶完整性、氨基酸序列相似度及遗传距离, 从美洲大蠊基因组鉴定出一类 BovB 逆转录转座子: RTE-7_PAm。BovBs 的反转录酶氨基酸序列的系统树表明, 美洲大蠊与内华达古白蚁 *Zootermopsis nevadensis* 的进化关系比与其同属蜚蠊科 Blattidae 的德国小蠊的关系更近。昆虫中 BovB 逆转录转座子的进化关系与传统的核基因进化关系的不同, 表明转座子的进化相对宿主基因的进化具有一定的独立性。

关键词: 美洲大蠊; 基因组; 重复序列; 转座元件; BovB 逆转录转座子; 系统进化

中图分类号: Q78

文献标志码: A

Analysis of Repetitive Sequences in *Periplaneta americana* Genome

MOU Biqin¹, YAN Chaochao¹, LI Wujiao¹, LI Jing¹, SHEN Yongmei², YUE Bisong^{1*}

(1. Key Laboratory of Bio-Resources and Eco-Environment, Ministry of Education, College of Life and Sciences, Sichuan University, Chengdu 610064, China; 2. Sichuan Key Laboratory of Medicinal *Periplaneta Americana*, Chengdu 610081, China)

Abstract: Repetitive sequences constitute a large fraction of a eukaryote genome. Some types of repetitive sequences, such as LINEs, which are a member of autonomous transposons, have been widely applied in molecular phylogenetic and genetic diversity studies of insects. In this study, *de novo* prediction and homology alignment were used to search the repetitive sequences in *Periplaneta americana* genome, and the results showed that the screened repetitive sequences accounted for approximately 62% of the genome. Among interspersed repetitive sequences, the DNA transposons constituted 16.18% of the whole genome. In retrotransposons, LINEs accounted for 13.64% of the genome followed by SINE (3.52%) and LTR (1.32%), respectively. The BovB superfamily in LINEs was the most abundant (6.73%) compared to other types of transposons. The proportion of all types of transposable element in *P. americana* were higher than that of *Blattella germanica* except the LTR retrotransposons. According to the integrality of the reverse transcriptase of retrotransposons, similarities and genetic distance of the amino acid sequences, a kind of BovB retrotransposons were identified in *P. americana* genome and named RTE-1_Pam. Phylogenetic analyses based on the amino acid sequences of BovB reverse transcriptase indicated that *P. americana* shared closer relationship with *Zootermopsis nevadensis* than with *B. germanica* which are both Blattidae. The difference of phylogenetic analyses results among these insects inferred by using retrotransposons and karyogene suggested an independent evolutionary history of TEs relative to the host karyogene.

Keywords: *Periplaneta americana*; genome; repetitive sequences; transposable elements; BovB retrotransposons; phylogenetic

重复序列是真核基因组的重要组成部分, 它分为串联重复序列和散在重复序列, 其中转座子是散在重复序列的主要组成部分。转座子又称为转座元件(transposable element, TE), 是广泛分布于真核生物基因组中的一种可移动的 DNA 序列(Kidwell & Lisch, 2002), 是宿主进化的重要动力之一(Kazazian, 2004)。按转座机制和序列特征, 散在重复序列可分为 DNA 转座子和逆转录转座子两大类。逆转录转座子编码逆转录酶, 该酶将由 DNA 转录形成的 RNA 经逆转录过程转换成 cDNA, 最后将形成的 cDNA 整合到基因组里完成转座过程, 引起稳定突变(Eickbush *et al.*, 2002)。根据序列两端是否含有长末端重复序列, 逆转录转座子又分为长末端重复序列(LTR)和非长末端重复序列。非自主的非长末端重复序列称为短散在重复序列(SINE)。非自主重复序列不能独立转座, 自身不编码有功能的逆转录转座酶, 只能依赖于相关的自主元件进行转座。相反的, 自主的非长末端重复序列称为长散在重复序列(LINE),

收稿日期: 2017-02-22

接受日期: 2017-04-25

作者简介: 牟必琴, 女, 硕士研究生, 研究方向: 基因和遗传信息分析, E-mail:435077636@qq.com

*通信作者 Corresponding author, E-mail:bsyue@scu.edu.cn

它本身含有有功能的反转录酶而完成自身的转座过程。DNA 转座子则编码转座酶，一般直接通过“剪切和粘贴”(cut and paste)机制发生转座，引起不稳定突变(Ohshima *et al.*, 1996)。

关于昆虫 BovB 逆转录转座子(BovBs)的进化研究迄今还比较少。BovBs 属于 LINE，全长约 3.2 kb，在蜚蠊目 Blattodea 物种的基因组中均占很高比例。完整的 BovBs 中间含有一个长约 1 000 个氨基酸的开放阅读框(ORF)，用于编码逆转录转座子转座活动过程中需要的 4 种酶：反转录酶、核酸内切酶、核酸外切酶及磷酸酶(Malik & Eickbush, 1998)。BovBs 在各类真核生物的基因组中均有报道，比如在短尾负鼠 *Monodelphis domestica*、牛、蜥形类和猛犸象 *Mammuthus primigenius* 等物种的基因组中被发现占很高比例(Gentles *et al.*, 2007; Adelson *et al.*, 2009; Kordis, 2009; Zhao & Ji, 2009)。近年来，昆虫中的 BovBs 也陆续地被发现：如鳞翅目 Lepidoptera 的家蚕 *Bombyx mori*(Morton *et al.*, 2012)和肩突硬蜱 *Ixodes scapularis*(Tay *et al.*, 2010)。此外，一些研究对蝇蛹金小峰 *Nasonia vitripennis*、红带釉蝶 *Heliconius Melpomene*、果蝇类 *Drosophila*、埃及伊蚊 *Aedes aegypti*、致倦库蚊 *Culex quinquefasciatus* 和按蚊类 *Anopheles* 基因组中 BovBs 进行了结构和活性的相关研究(Clark & Eisen, 2007; Sinkins, 2007; Arensburger *et al.*, 2010; Webb *et al.*, 2011; Consortium *et al.*, 2012)。

随着二代测序技术的飞速发展，越来越多物种的全基因组得以揭示，在物种的全基因组中鉴定重复序列的研究也越来越多。美洲大蠊 *Periplaneta americana* 属于蜚蠊目蜚蠊科 Blattellidae 大蠊属 *Periplaneta*，具有较强的抗逆和适应能力，是世界性的卫生害虫，同时在中国也是重要的药用昆虫。蜚蠊目的德国小蠊 *Blattella germanica* 为其同目同科物种。近年来，原等翅目的内华达古白蚁 *Zootermopsis nevadensis* 被归类于蜚蠊目(Terrapon *et al.*, 2014)。为了了解美洲大蠊的遗传背景，我们结合 *de novo* 预测和同源性搜索两种相结合的方法首次对美洲大蠊基因组重复序列进行鉴定，并比较分析了重复序列在美洲大蠊、德国小蠊和内华达古白蚁 3 种蜚蠊目昆虫基因组中的分布规律，最后以 BovB 逆转录转座子内反转录酶编码序列作为分子标记，初步研究了 BovB 逆转录转座子在 16 种昆虫中的进化关系。

1 基因组来源与方法

1.1 基因组来源

美洲大蠊样品来自于四川好医生攀西药业有限公司西昌饲养基地。使用二代 Illumina 和三代 Pacbio 测序相结合的方法，分别得到了 460X 和 10X 的原始数据。通过软件 DISCOVAR (<https://www.broadinstitute.org/software/discovar/blog>) (Love *et al.*, 2016)和 BESST(Sahlin *et al.*, 2014, 2016)组装得到美洲大蠊全基因组。德国小蠊和古白蚁的基因组均下载自 NCBI (PRJNA203136 和 PRJNA203242)。

1.2 方法

1.2.1 重复序列的鉴定 使用 3 种方案搜集预测基因组中的重复序列：(1)从头预测转座子：使用 RepeatModeler 软件(<http://www.repeatmasker.org>, version 1.0.8)扫描基因组序列，对不同重复序列的特性进行计算机建模，得到非冗余的重复序列数据库(Smit *et al.*, 2014b)，然后使用 RepeatMasker 软件(<http://www.repeatmasker.org>, version 4.0.5)进行重复序列的搜索和识别(Smit *et al.*, 2014a)。(2)基于同源比对注释重复序列：使用 RepeatMasker 软件和 RepBase 重复序列数据库 (version 19.09)进行比对并注释(Jurka *et al.*, 2005; Bao *et al.*, 2015)。(3)使用软件 RepeatProtein(version 4.0.5)和它的默认参数搜索已知转座子内的蛋白质序列相关的重复序列(Hu *et al.*, 2012)。

1.2.2 重复序列的统计 重复序列统计的原始文件为 RepeatMasker 的输出文件(“.out”)和 RepeatProtein 的输出文件(“.annot”)。对重复序列进行统计之前，我们首先对每种类型的重复序列进行过滤，去掉核苷酸序列长度小于 10 bp 和与重复序列库的相似度小于 70%的重复元件。这一过程使用 shell 和 python 编写的本地脚本完成。过滤得到的文件使用 R 语言统计每个物种基因组中重复序列的种类、占基因组的比例和作图。

1.2.3 BovB 逆转录转座子进化分析 使用 LINE 保守的反转录酶区域的氨基酸序列构建昆虫的系统进化树。首先，将参考基因组的 BovB 逆转录转座子序列作为质询序列，使用 blastx 搜索出与 NCBI 中的蛋白保守序列库(CDD)中 LINE 中的反转录酶氨基酸保守序列(cd01650)同源性较高的蛋白质序列(evalue 值不大于 1e-10, 氨基酸长度大于 100 bp)(Sun & Mueller, 2014);然后将上述所得的蛋白质序列进行修整(如去除低质量的插入缺失位点)使保留 CDD 库内的一致性序列，然后使用 CD-HIT 去除冗余序列(Huang *et al.*, 2010)，得到氨基酸序列；最后使用 RepeatMasker 中的 COSEG(Lavoie *et al.*, 2013)合并得到 BovBs 中氨基酸的一致性序列。使用 MEGA 5.2(Tamura *et al.*, 2011; Sormacheva *et al.*, 2012)计算氨基酸序列之间的遗传距离。此外，还从 RepBase 重复序列蛋白质序列库中下载了其他 13 种昆虫的 19 条 BovBs 蛋白质序列(表 1)。使用 ClustalW (Thompson *et al.*, 1999)和 prank v.140603 (Löytynoja & Goldman, 2010)将氨基酸序列集进行对齐。最后使用 MEGA 5.2(Tamura *et al.*, 2011)构建最大似然树，Bootstrap 设置为 1 000 以检验分子系统树各分支的置信度，并使用 MrBayes 3.2(Ronquist & Huelsenbeck, 2003)构建贝叶斯树。

表 1 BovB 逆转录转座子氨基酸序列信息

Table 1 The amino acid sequence information of BovB retrotransposons

物种拉丁名	物种中文名	目	BovB 逆转录转座子建树序列命名
<i>Periplaneta americana</i>	美洲大蠊	蜚蠊目 Blattodea	Paa_RTE-1_PAm
<i>Blattella germanica</i>	德国小蠊	蜚蠊目 Blattodea	Bga_RTE-1_BGe
<i>Zootermopsis nevadensis</i>	内华达古白蚁	蜚蠊目 Blattodea	Zns_RTE-1_ZNe
<i>Nasonia vitripennis</i>	蝇蛹金小蜂	膜翅目 Hymenoptera	Nvs_RTE-1_NVi
<i>Bombyx mori</i>	家蚕	鳞翅目 Lepidoptera	Bmi_RTE-1_BM
<i>Heliconius Melpomene</i>	红带釉蝶	鳞翅目 Lepidoptera	Hme_RTE-1_HMe
<i>Anopheles gambiae</i>	冈比亚按蚊	双翅目 Diptera	Age_RTE-3_AG
<i>Culex quinquefasciatus</i>	致倦库蚊	双翅目 Diptera	Cqs_RTE-2_CQ
<i>Culex quinquefasciatus</i>	致倦库蚊	双翅目 Diptera	Cqs_RTE-3_CQ
<i>Aedes aegypti</i>	埃及伊蚊	双翅目 Diptera	Aai_RTE_Ele4
<i>Anopheles genus</i>	按蚊属	双翅目 Diptera	Ags_RTE-1_AG
<i>Drosophila yakuba</i>	yakuba 果蝇	双翅目 Diptera	Dya_RTE-1_DYa
<i>Drosophila ficusphila</i>	ficusphila 果蝇	双翅目 Diptera	Dfa_RTE-1_DF
<i>Drosophila ficusphila</i>	ficusphila 果蝇	双翅目 Diptera	Dfa_RTE-2_DF
<i>Drosophila ficusphila</i>	ficusphila 果蝇	双翅目 Diptera	Dfa_RTE-3_DF
<i>Drosophila ficusphila</i>	ficusphila 果蝇	双翅目 Diptera	Dfa_RTE-4_DF
<i>Drosophila ananassae</i>	嗜凤梨果蝇	双翅目 Diptera	Dae_RTE-1_DAn
<i>Drosophila ananassae</i>	嗜凤梨果蝇	双翅目 Diptera	Dae_RTE-2_DAn
<i>Drosophila bipectinata</i>	双栉形果蝇	双翅目 Diptera	DbA_RTE-1_DBP
<i>Drosophila bipectinata</i>	双栉形果蝇	双翅目 Diptera	DbA_RTE-2_DBP
<i>Drosophila biarmipes</i>	biarmipes 果蝇	双翅目 Diptera	DBs_RTE-1_DBi
<i>Drosophila takahashii</i>	高桥果蝇	双翅目 Diptera	Dti_RTE-1_DT

注: BovBs 序列名称的命名规则为: 物种拉丁名属名的首字母和种名首尾字母+BovB 逆转录转座子的亚家族名。

Notes: The sequence names consists of the first letter of family' Latin name, the first and the end letters of species' name as well as the subfamily name of BovB retrotransposons.

2 结果

2.1 主要的重复序列和占基因组比例

在美洲大蠊基因组中, 使用 RepeatModeler *de novo* 预测方法贡献最大, 得到了 1 862.3 Mp 的重复序列, 占基因组的 58.27%; 使用 RepBase 同源比对注释得到了 355.3 Mp 的重复序列, 占基因组的 11.12%; 使用 RepeatProtein 软件识别的转座子蛋白序列库得到了 417.5 Mp 的重复序列, 占基因组的 13.06%。综合以上 3 种不同的方法, 去除冗余后(不同方法可能在同一个位置预测出转座子), 我们总共得到了 1 949.5 Mp 的非冗余重复序列, 占基因组的 62.38%(表 2, 图 1), 这和人类基因组的重复序列占比相似(International Human Genome Sequencing Consortium, 2001)。在美洲大蠊基因组重复序列中, 未知类型的重复序列(Unknown)占总重复序列的 1/3 左右, 占基因组的 25.75%。已知类型转座子中 DNA 转座子总长度占基因组的 16.18%, 相对于所有的逆转录转座子(占基因组比例为 18.75%)相对较小。其中 LINE 逆转录转座子为 13.64%。此外, SINE 和 LTR 逆转录转座子比例均较少, 只占大约 1.80%左右。其他类型的重复序列, 如卫星重复序列、核糖体重复序列及其他简单重复序列占基因组的 3.70%(表 2, 图 1)。

蜚蠊目物种中不同类型的重复序列比较分析结果见表 3。美洲大蠊基因组的重复序列含量(62.38%)略高于德国小蠊基因组重复序列的含量(61.94%), 远高于内华达古白蚁(35.84%)。这一结果与脊椎动物中随着基因组的增加重复序列占基因组比例增大的趋势一致(Chalopin *et al*, 2014; Chalopin *et al*, 2015)。DNA 型重复序列是美洲大蠊和德国小蠊基因组中最丰富的重复序列类型, 其中美洲大蠊的 DNA 型重复序列含量最高, 达 16.18%, 德国小蠊次之(13.07%), 均显著高于内华达古白蚁(6.20%)。这一现象也出现在同属于 DNA 重复序列的 RC/Helitron 型重复序列中: 美洲大蠊最高, 内华达古白蚁最低。而 LINE 逆转录转座子是古白蚁基因组中最丰富的重复序列类型, 约占基因组 10%, 但这一比例仍然低于美洲大蠊(13.64%)和德国小蠊(12.13%)。SINE 重复序列在美洲大蠊和德国小蠊中的含量大致相当(大约 3%), 并且均高于内华达古白蚁(2.24%)。虽然大多数转座子类型在美洲大蠊基因组中都超过德国小蠊, 但德国小蠊基因组中 LTR 转座子(5.02%)的比例却显著高于美洲大蠊(1.78%)。

表 2 美洲大蠊基因组重复序列的类型和比较分析

Table 2 Comparison of each class of repetitive sequences in *Periplaneta americana* genome

重复序列种类 Classes of repeats	RepeatModeler 注释 RepeatModeler		RepBase 库注释 RepBase		RepeatProtein 库注释 RepeatProtein		重复序列总和 Total	
	长度 Length/Mp	占基因组比例 Percentage of genome/%	长度 Length/Mp	占基因组比例 Percentage of genome/%	长度 Length/Mp	占基因组比例 Percentage of genome/%	长度 Length/Mp	占基因组比例 Percentage of genome/%
	DNA	442.4	13.84	112.9	3.53	141.9	4.44	517
LINE	327.1	10.23	123.8	3.87	268.5	8.4	436	13.64
SINE	112.3	3.52	9.5	0.3	0	0	113.9	3.56
LTR	42.2	1.32	13.3	0.42	5.7	0.18	57	1.78
RC/Helitron	57.6	1.8	4.1	0.13	1.8	0.06	59.1	1.85
Other	86.8	2.72	99.8	3.12	0	0	118.2	3.7
Unknown	822.2	25.72	0.9	0.03	0	0	823.1	25.75
total	1 862.3	58.27	355.3	11.12	417.5	13.06	1 993.9	62.38

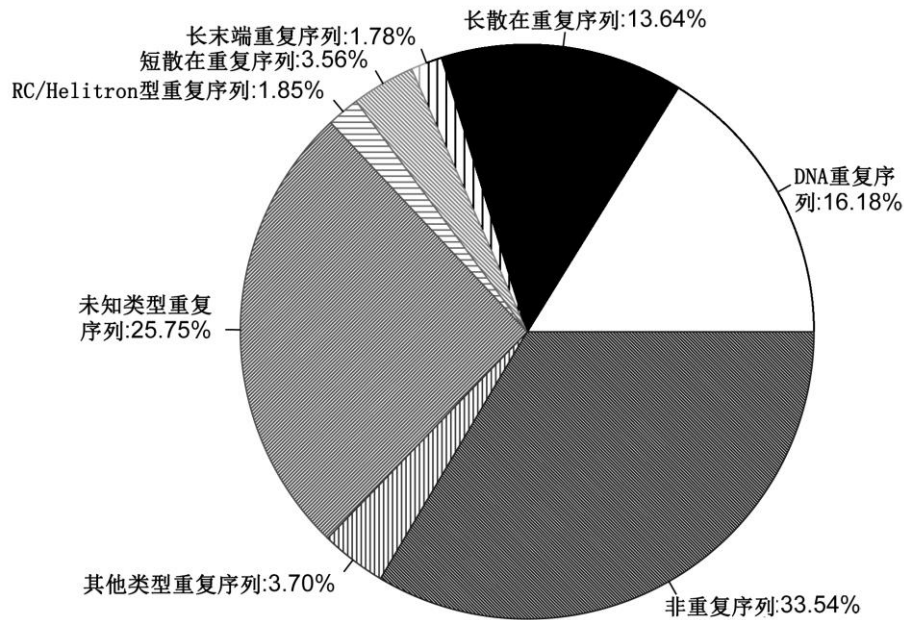


图 1 美洲大蠊基因组中重复序列分布图

Fig. 1 The pie distribution diagram of the repetitive sequences in the *American cockroach* genome

表 3 不同物种重复序列的比较

Table 3 Comparison of repetitive sequences in different species

重复序列种类 Classes of repeats	美洲大蠊		德国小蠊		内华达古白蚁	
	长度/Mp	占基因组比例/%	长度/Mp	占基因组比例/%	长度/Mp	占基因组比例/%
	Length/Mp	Percentage of genome/%	Length/Mp	Percentage of genome/%	Length/Mp	Percentage of genome/%
DNA	517	16.18	223.5	13.07	28.8	6.2
LINE	436	13.64	207.5	12.13	46.4	9.98
SINE	113.9	3.56	52.8	3.07	10.4	2.24
LTR	57	1.78	85.8	5.02	26.5	5.7
RC/Helitron	59.1	1.85	3.9	0.23	1.5	0.32
Other	118.2	3.7	46.6	2.73	6	1.29
Unknown	823.1	25.75	569.9	33.32	66.4	14.31
total	1 993.9	62.38	1059.3	61.94	1 666.4	35.84

2.2 主要重复序列家族比较分析

在所有的重复序列类型当中，DNA 转座子亚家族(TcMar、hAT)和 LINE 逆转录转座子亚家族(BovB、L2)占 3 个基因组比例最高(图 2)，尤其是 BovB 逆转录转座子在本研究中 3 个近缘物种的比例均最高，美洲大蠊的 BovBs 也高于其他 2 个物种。而 LTR 逆转录转座子亚家族中，德国小蠊的 Copia、Gypsy 和 Pao 亚家族比例高于美洲大蠊。相反，美洲大蠊的 SINE 类型的 ID 亚家族比例明显高于德国小蠊的。除此之外，如图 2 所示，内华达古白蚁的 BovBs 在其基因组中的比例最高(5.92%)，且高于德国小蠊(4.50%)，这归功于 BovB 属于新型转座子类型的 LINE 逆座子。

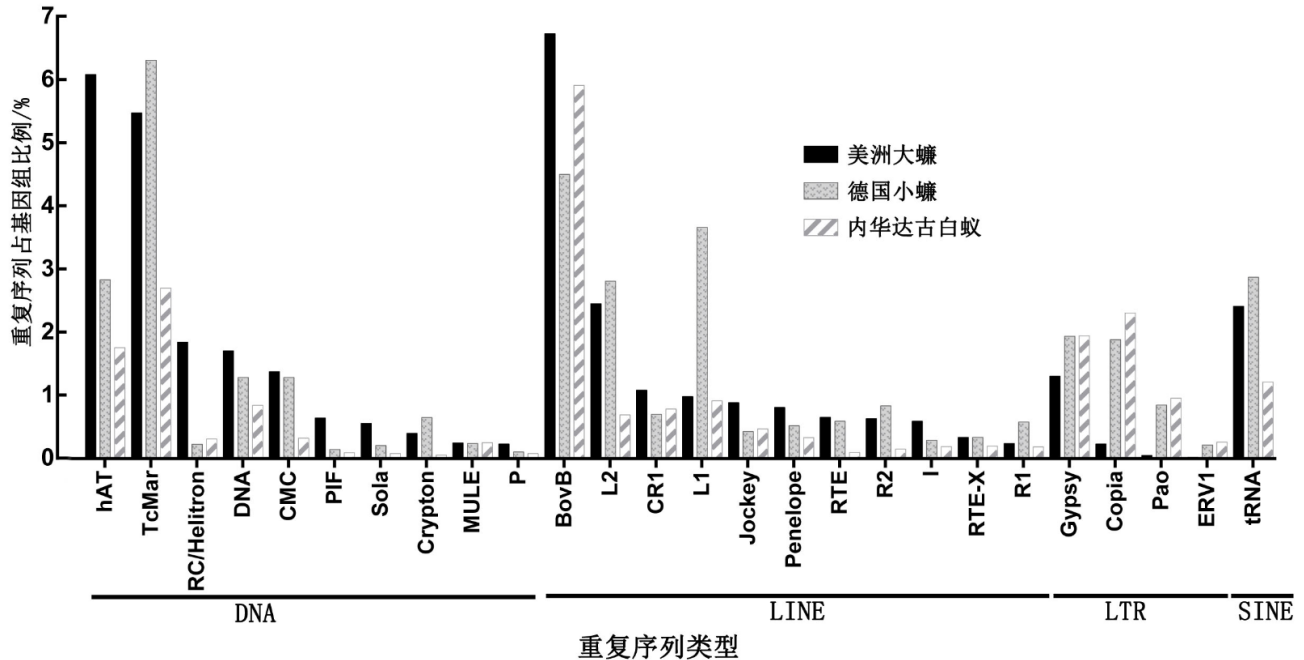


图 2 3 种蜚蠊目昆虫基因组主要重复序列亚家族比较分析

Fig. 2 Comparative analysis of the major repetitive sequence subfamilies in three Blattodea insect genomes

2.3 美洲大蠊 BovB 逆转录转座子的序列分析

由于 BovBs 在美洲大蠊、德国小蠊及内华达古白蚁中占基因组比例均很高，可被用于转座子系统进化研究。将 BovBs 通过 blastx 比对到反转录酶氨基酸序列库，分别得到了美洲大蠊、德国小蠊和内华达古白蚁具有 90% 以上同源性的一系列氨基酸序列。使用 COSEG 得到它们的一致性序列并分别命名为 RTE-1_PAm, RTE-1_BGe 和 RTE-1_ZNe。将它们与其他 13 种昆虫的 19 条 BovBs 氨基酸序列一起进行比较分析。本研究的蜚蠊目昆虫和部分其他目昆虫的遗传距离及相似性结果见表 4。结果表明蜚蠊目和与蜚蠊目相关的其他昆虫 BovBs 反转录酶的遗传距离介于 0.29~1.14，序列的相似度介于 34.3%~75.2%。其中美洲大蠊与内华达古白蚁的反转录酶具有最高的相似度，为 75.2%，与之对应，两者遗传距离最小，为 0.29。美洲大蠊与德国小蠊的反转录酶的相似度为 65.2%，它们之间的遗传距离值为 0.43，明显地大于美洲大蠊与内华达古白蚁之间的遗传距离值。美洲大蠊与红袖粘蝶的 BovBs 反转录酶同源性最低，相似度为 34.3%，遗传距离值为 1.07。

表 4 昆虫 BovB 逆转录转座子氨基酸序列的遗传距离及相似性(%)

Table 4 The genetic distance and similarity (%) of BovB retrotransposons in insect genomes

	美洲大蠊	德国小蠊	内华达古白蚁	红带粘蝶	蝇蛹金小蜂
美洲大蠊		65.2	75.2	34.3	60.2
德国小蠊	0.43		67.3	35.0	57.8
内华达古白蚁	0.29	0.40		36.9	61.7
红带粘蝶	1.07	1.05	0.99		31.9
蝇蛹金小蜂	0.51	0.55	0.48	1.14	

注：对角线上半部分数据为相似度，下半部分数据为遗传距离

Notes: The data above the diagonal are the similarities between BovB retrotransposons, while those below the diagonal are the genetic distance.

2.4 BovB 逆转录转座子系统进化分析

将 RTE-1_PAm, RTE-1_BGe 和 RTE-1_ZNe 与其他昆虫的 19 条 BovBs 家族转座子的氨基酸进行序列比对, 采用贝叶斯法和最大似然法构建分子系统发育树。来自 4 个目的 16 个物种的 22 条氨基酸序列被分为 4 大类群。I 类群包含来自鳞翅目的家蚕和红带袖蝶(图 3)。类群 II 和 III 均属于双翅目, 类群 II 含有双翅目的果蝇属物种, 与鳞翅目聚在一起。类群 III 既含有果蝇类, 又含有蚊类, 其中按蚊属与果蝇属聚在一个小支上。其他 3 种蚊类聚集在另外一个小支上, 包括 1 种埃及伊蚊、2 条致倦库蚊和 1 种冈比亚按蚊。类群 IV 中包括膜翅目的蝇蛹金小蜂和蜚蠊目的美洲大蠊、德国小蠊以及内华达古白蚁, IV 位于系统进化树内部, 证明我们鉴定的蜚蠊目的 BovBs 非常可靠。美洲大蠊的 BovBs 和内华达古白蚁的 BovBs 聚集在同一个进化小支上, 且两者的进化距离较德国小蠊与美洲大蠊的距离更近(图 3, 表 4)。

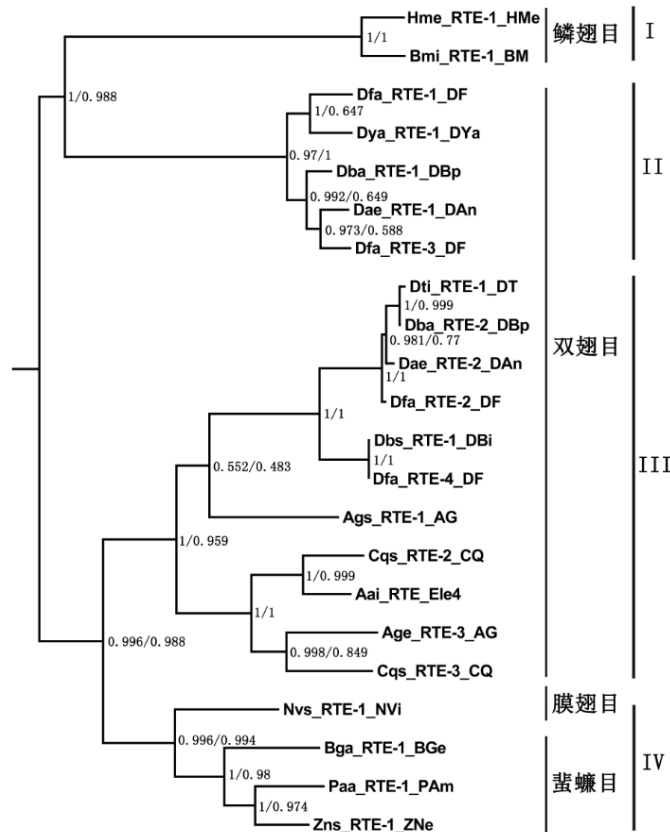


图 3 BovB 逆转录转座子反转录酶氨基酸序列系统进化树

Fig. 3 Phylogenetic relationship among BovB retrotransposons based on the amino acid sequence of reverse transcriptase

基于 BovB 逆转录转座子的 15 种昆虫系统进化关系。贝叶斯推断和最大似然法树。枝上的数字分别表示贝叶斯后验概率和最大似然法 1 000 次 Bootstrap 检验置信度。

Phylogenetic tree of BovB retrotransposon that constructed by Bayesian Inference and Maximum likelihood methods. The numbers beside the nodes are Bayesian posterior probabilities and ML bootstrap proportions.

3 讨论

在 DNA 转座子类型中, 两种蜚蠊科昆虫的 DNA 转座子占基因组的比例均远高于内华达古白蚁, 这种差异主要表现在两种蜚蠊科昆虫 DNA 转座子中的 2 个超家族 hAT 和 TcMar 占基因组的比例明显高于内华达古白蚁。在已知的反转录转座子类型中, 美洲大蠊 LTR 的类型和占基因组的比例明显低于德国小蠊, 但其他转座子类型均是美洲大蠊略高于德国小蠊。从总体上看, 美洲大蠊、德国小蠊和内华达古白蚁重复序列占基因组比例呈递减趋势, 这与它们基因组大小依次递减一致, 符合随着基因组的增大, 重复序列占基因组比例也增大的规律(Chalopin *et al.*, 2014; Chalopin *et al.*, 2015)。美洲大蠊重复序列最多的类型为 Unknown(25.75%), 说明其基因组中未知的重复序列类型还有很多, 在这些未知类型的重复序列中除一部分简单重复序列外, 还可能存在一些物种特异的新类型重复序列, 有必要进一步挖掘分析。

采用存在于昆虫基因组中且某些物种含量丰富的逆转录转座子 BovBs 进行了比较研究和系统进化分析, 为理解 BovB 逆转录转座子在昆虫基因组的组成和进化历程奠定了一定的基础。美洲大蠊、德国小蠊和内华达古白蚁三个

物种基因组中的一类 BovBs 序列，它们的氨基酸同源性高达 90%。比对发现，RTE-1_PAm 类元件与的 RTE-1_BGe 相似度为 65.2%，而与内华达古白蚁 RTE-1_ZN 聚在同一支上且其相似度为 75.2%，与传统的分类和系统进化关系(德国小蠊与美洲大蠊关系较近)不同(Misof *et al.*, 2014)。形成这种差异的原因可能是转座子的进化速率与基因组的进化速率的不一致，即转座子的同义替换速率要低于或高于宿主基因的同义替换速率(Sánchezgracia *et al.*, 2005)。除此之外，选择压力和修复机制的不同也有可能使转座子与宿主基因的进化速率出现差异。要解决这一问题，需要对更多昆虫基因组的特性和转座子的进化历程进行更深入的研究。

转座子在基因组中随意跳动，导致基因组序列的删除、扩增、移位、断裂和重组等现象(Langley *et al.*, 1988; Charlesworth *et al.*, 1994)，从而影响物种的正常生命活动(Le & Cappy, 2006; Oliver & Greene, 2009)。近年来，昆虫中重复序列的研究较少，已有的研究主要集中于鳞翅目和双翅目的果蝇属当中。例如家蚕基因组中的 MITEs、Chapaev、Tc1/Mariner 等转座子的转座活动使基因组发生重组(Zhang *et al.*, 2014)。果蝇属的 P 转座子的扩增和水平转移影响了基因组的变化(Clark & Kidwell, 1997; Arcà & Savakis, 2000)。本研究首次对美洲大蠊基因组的重复序列进行了系统的统计描述并结合近缘物种的重复序列进行比较分析，这为美洲大蠊基因组研究提供了重要支持，对昆虫重复序列的研究具有积极作用。

参考文献

- Adelson DL, Raison JM, Edgar RC. 2009. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome[J]. *Proceedings of the National Academy of Sciences*, 106(31): 12855-12860.
- Arcà B, Savakis C. 2000. Distribution of the transposable element Minos in the genus *Drosophila*[J]. *Genetica*, 108(3): 263.
- Arensburger P, Megy K, Waterhouse RM, *et al.* 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics[J]. *Science*, 330(6000): 86-88.
- Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes[J]. *Mobile DNA*, 6(1): 11.
- Chalopin D, Fan S, Simakov O, *et al.* 2014. Evolutionary active transposable elements in the genome of the coelacanth[J]. *Journal of Experimental Zoology Part B Molecular & Developmental Evolution*, 322(6): 322-333.
- Chalopin D, Naville M, Plard F, *et al.* 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates[J]. *Genome Biology Evolution*, 7(2): 567-580.
- Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes[J]. *Nature*, 371(6494): 215-220.
- Clark AG, Eisen MB. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny[J]. *Nature*, 450(7167): 203-218.
- Clark JB, Kidwell MG. 1997. A phylogenetic perspective on P transposable element evolution in *Drosophila*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 94(21): 11428.
- Consortium THG, Dasmahapatra KK, Walters, *et al.* C. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species[J]. *Nature*, 487(7405): 94-98.
- Eickbush TH, Malik HS, Eickbush TH. 2002. Origins and Evolution of Retrotransposons[J]. *Mobile DNA*: 1111-1144.
- Gentles AJ, Wakefield MJ, Kohany O, *et al.* 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*[J]. *Genome Research*, 17(7): 992-1004.
- Hu Q, Ma T, Wang K, *et al.* 2012. The Yak genome database an integrative database for studying yak biology and high-altitude adaption[J]. *BMC Genomics*, 13(2): 600.
- Huang Y, Niu B, Gao Y, *et al.* 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences[J]. *Bioinformatics*, 26(5): 680-682.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome[J]. *Nature*, 409: 860-921.
- Jurka J, Kapitonov VV, Pavlicek A, *et al.* 2005. Repbase Update, a database of eukaryotic repetitive elements[J]. *Cytogenetic & Genome Research*, 110(1-4): 462.
- Kazazian HH. 2004. Mobile elements: drivers of genome evolution[J]. *Science*, 303(5664): 1626-1632.
- Kidwell M, Lisch D, 2002. Transposable elements as sources of genomic variation[M]// In: Craig NL, Cragie R, Gellert M, *Lambowitz A (ed).*, *Mobile DNA II*. Washington DC: ASM Press, 59-90.
- Kordis D. 2009. Transposable elements in reptilian and avian (sauropsida) genomes[J]. *Cytogenetic & Genome Research*, 127(2-4): 94-111.
- Langley CH, Montgomery E, Hudson R, *et al.* 1988. On the role of unequal exchange in the containment of transposable element copy number[J]. *Genetics Research*, 52(3): 223-235.
- Lavoie CA, Platt RN, Novick PA, *et al.* 2013. Transposable element evolution in *Heliconius*, suggests genome diversity within Lepidoptera[J].

- Mobile DNA, 4(1):21.
- Le RA, Capy P. 2006. Population genetics models of competition between transposable element subfamilies[J]. *Genetics*, 174(2): 785-793.
- Love RR, Weisenfeld NI, Jaffe DB, *et al.* 2016. Evaluation of DISCOVAR *de novo*, using a mosquito sample for cost-effective short-read genome assembly[J]. *BMC Genomics*, 17(1): 1-10.
- Löytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser[J]. *BMC Bioinformatics*, 11(1): 1-6.
- Malik HS, Eickbush TH. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs[J]. *Molecular Biology and Evolution*, 15(9): 1123-1134.
- Misof B, Liu S, Meusemann K, *et al.* 2014. Phylogenomics resolves the timing and pattern of insect evolution[J]. *Science*, 346(346): 763-767.
- Morton WA, Kortschak RD, Gardner MG, *et al.* 2012. Widespread horizontal transfer of retrotransposons[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 110(3): 1012-1016.
- Ohshima K, Hamada M, Terai Y, *et al.* 1996. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements[J]. *Molecular & Cellular Biology*, 16(7): 3756-3764.
- Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution[J]. *Bioessays*, 31(7): 703-714.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models[J]. *Bioinformatics*, 19(12): 1572-1574.
- Sahlin K, Chikhi R, Arvestad L. 2016. Assembly scaffolding with PE-contaminated mate-pair libraries[J]. *Bioinformatics*, 32(13): btw064.
- Sahlin K, Vezzi F, Nystedt B, *et al.* 2014. BESST-Efficient scaffolding of large fragmented assemblies[J]. *BMC Bioinformatics*, 15(1): 281.
- Sánchezgracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*[J]. *Trends in Genetics*, 21(4): 200.
- Sinkins S. 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector[J]. *Science*, 316(5832): 1718-1723.
- Smit A, Hubley R, Green P 2014b. RepeatModeler Open-1.0. 2008-2010[M].
- Smit A, Hubley R, Green P. 2014a. RepeatMasker Open-4.0. 2013–2015 [J]. URL <http://www.repeatmasker.org>.
- Sormacheva I, Smyshlyaev G, Mayorov V, *et al.* 2012. Vertical evolution and horizontal transfer of CR1 non-LTR retrotransposons and Tc1/mariner DNA transposons in Lepidoptera species[J]. *Molecular Biology & Evolution*, 29(12): 3685-3702.
- Sun C, Mueller RL. 2014. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders[J]. *Genome Biology & Evolution*, 6(7): 1818-1829.
- Tamura K, Peterson D, Peterson N, *et al.* 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. *Molecular Biology & Evolution*, 28(10): 2731-2739.
- Tay WT, Behere GT, Batterham P, *et al.* 2010. Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes[J]. *BMC Evolutionary Biology*, 10(1): 144.
- Terrapon N, Li C, Robertson HM, *et al.* 2014. Molecular traces of alternative social organization in a termite genome[J]. *Nature Communications*, 5(6183): 3636.
- Thompson JD. 1999. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice[J]. *Nucleic Acids Research*, 22(22): 4673-4680.
- Webb CHT, Riccitelli NJ, Ruminski DJ, *et al.* 2011. Widespread occurrence of self-cleaving ribozymes[J]. *Science*, 326(5955): 953.
- Zhang HH, Feschotte C, Han MJ, *et al.* 2014. Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals[J]. *Genome Biology & Evolution*, 6(6): 1375.
- Zhao FQ, Qi J, Schuster SC. 2009. Tracking the past: interspersed repeats in an extinct Afrotherian mammal, *Mammuthus primigenius*[J]. *Genome Research*, 19(19): 1384-1392.