NRC Research Press

ARTICLE

Random DNA libraries from three species of the stick insect genus *Bacillus* (Insecta: Phasmida): repetitive DNA characterization and first observation of polyneopteran MITEs

Marco Ricci, Andrea Luchetti, Livia Bonandin, and Barbara Mantovani

Abstract: The repetitive DNA content of the stick insect species *Bacillus rossius* (facultative parthenogenetic), *Bacillus grandii* (gonochoric), and *Bacillus atticus* (obligate parthenogenetic) was analyzed through the survey of random genomic libraries roughly corresponding to 0.006% of the genome. By repeat masking, 19 families of transposable elements were identified (two LTR and six non-LTR retrotransposons; 11 DNA transposons). Moreover, a de novo analysis revealed, among the three libraries, the first MITE family observed in polyneopteran genomes. On the whole, transposable element abundance represented 23.3% of the genome in *B. rossius*, 22.9% in *B. atticus*, and 18% in *B. grandii*. Tandem repeat content in the three libraries is much lower: 1.32%, 0.64%, and 1.86% in *B. rossius*, *B. grandii*, and *B. atticus*, respectively. Microsatellites are the most abundant in all species. Minisatellites were only found in *B. rossius* and *B. atticus*, and five monomers belonging to the *Bag320* satellite family were detected in *B. atticus*. Assuming the survey provides adequate representation of the relative genome, the obligate parthenogenetic species (*B. atticus*), compared with the other two species analyzed, does not show a lower transposable element content, as expected from some theoretical and empirical studies.

Key words: genomic sequence survey, miniature inverted repeats (MITEs), stick insects, tandem repeats, transposable elements.

Résumé : Les auteurs ont analysé le contenu en ADN répété chez les phasmes *Bacillus rossius* (à parthénogenèse facultative), *Bacillus grandii* (gonochorique) et *Bacillus atticus* (à parthénogenèse obligatoire) en explorant aléatoirement des banques génomiques correspondant à environ 0,006 % du génome. En masquant les séquences répétées, 19 familles d'éléments transposables ont été identifiées (deux familles de rétrotransposons à LTR et six familles sans LTR; 11 familles de transposons à ADN). De plus, une analyse de novo a permis de découvrir, au sein des trois banques, la première famille d'éléments MITE au sein des polynéoptères. Globalement, l'abondance des éléments transposables représentait 23,3 % du génome chez le *B. rossius*, 22,9 % chez le *B. atticus* et 18 % chez le *B. grandii*. Les séquences répétées en tandem étaient beaucoup moins abondantes au sein des trois banques : 1,32 %, 0,64 % et 1,86 % respectivement chez le *B. rossius*, le *B. grandii* et le *B. atticus*. Les microsatellites étaient les plus abondants chez les trois espèces. Des minisatellites n'ont été observés que chez le *B. rossius* et le *B. atticus*, tandis que cinq monomères de la famille de satellites *Bag320* ont été détectés de manière unique chez le *B. atticus*. En supposant que cet échantillonnage du génome soit suffisamment représentatif de chaque génome, il s'avère que le génome de l'espèce à parthénogenèse obligatoire (*B. atticus*) ne présente pas un contenu en éléments transposables inférieur à celui des deux autres espèces, tel que le prédisaient certaines études théoriques et empiriques. [Traduit par la Rédaction]

Mots-clés : relevé de séquences génomiques, séquences répétées inversées miniatures (MITE), phasmes, répétitions en tandem, éléments transposables.

Introduction

A significant fraction of eukaryotic genomes harbours DNA sequences repeated either in tandem (head-to-tail arranged) or interspersed (Richard et al. 2008).

Tandem repeats are made by monomeric units, whose length is comprised between two and hundreds of base pairs (bp), organized in arrays ranging from few to thousands of units. They can be categorized into the following three main classes: microsatellites, minisatellites, and satellite DNAs (satDNAs). Although the three classes cannot be discriminated on the sole basis of unit length (Charlesworth et al. 1994; Richard et al. 2008), their approximate monomer length ranges can be considered as 2–10, 11–100, and >100 bp, respectively.

Interspersed repeats are, mainly, transposable elements (TEs), i.e., sequences able to move from one genomic location to another (Makałowski et al. 2012). There are two main classes of TEs: class I elements, moving via an RNA intermediate (retrotransposons), and class II elements, moving via a DNA intermediate (transposons). Within the two TE classes, autonomous elements, which are able to encode the proteins necessary for their transposition, and nonautonomous elements, which parasitize the transposition machinery of an autonomous partner, can be further distinguished (Makałowski et al. 2012). Among the latter, short interspersed elements (SINEs) are the most diverse and represented retrotransposons; their sequence is composed of (*i*) an RNA-related head, (*ii*) an anonymous body, and (*iii*) a simple sequence repeat tail. To date, the absence of SINEs has been reported only in species of *Drosophila* (Kramerov and Vassetzky 2011). Miniature inverted repeats (MITEs) are nonautonomous DNA transposons; they usually derive from autonomous elements through deletion of the internal protein-coding sequence and are characterized

Received 25 June 2013. Accepted 29 October 2013.

Corresponding Editor: T.E. Bureau.

M. Ricci, A. Luchetti, L. Bonandin, and B. Mantovani. Dipartimento di Scienze Biologiche, Geologiche e Ambientali (BiGeA), Università di Bologna, via Selmi 3, 40126 Bologna, Italy.

Corresponding author: Marco Ricci (e-mail: marco.ricci19@unibo.it).

Published at www.nrcresearchpress.com/gen on 7 November 2013.

by terminal or subterminal inverted repeats. MITEs are mainly found in plants, but they are also well represented in metazoan genomes (Feschotte et al. 2002).

Far from being just "junk DNA", repetitive DNA sequences are known to be biologically relevant for the host genome. SatDNAs, for example, are involved in structural functions, being the major component of centromeres in almost all eukaryotes (Plohl et al. 2008). On the other hand, TE genomic dynamics impact the host in different ways. For example, in many instances, TE insertions have been found to modify gene structures, gene expression profiles, and to promote recombination, thus introducing genetic diversity and even adaptive changes (Kazazian 2004). Besides positive interactions, repeated DNAs may also impact negatively on the host genome: for example, contraction or expansion of mini- and microsatellite arrays are directly involved in cancer development and other human pathologies (Galindo et al. 2011; El-Murr et al. 2012), while TE insertions can be deleterious for their involvement in gene disruption, negative alteration of gene expression, and ectopic recombination (Kazazian 2004).

The very existence of repeated DNAs has been long considered paradoxical because their accumulation should not be easily tolerated by the host genome (Charlesworth et al. 1994). In principle, TE accumulation or large expansions of tandem repeat arrays are contrasted by recombination that helps in eliminating deleterious alleles, for example, a TE insertion or a too large tandem array. Organisms with high genetic diversity and bisexual reproduction, therefore, would eliminate more efficiently an overload of repeated sequences. This relationship has been well depicted in the evolutionary hypothesis known as Muller's ratchet (Brookfield and Badge 1997; Wright and Schoen 1999): nonrecombining genomes would accumulate deleterious mutations that will drive them to extinction. However, the observation of unisexual and asexual taxa persisting over evolutionary time requires that mechanisms exist that are able to avoid the deleterious mutation load: the absence of TEs (or a very low TE activity) can be one such mechanism (Arkhipova and Meselson 2000, 2005; Sullender and Crease 2001). It is worth noting, though, that exceptions occur in that unisexual taxa show the same TE load than bisexual taxa (Kraaijeveld et al. 2012). Moreover, generally speaking, less virulent or even favourable parasites (here including TEs) can be selected in an unisexual or asexual lineage as a result of the strictly vertical transmission under these reproductive circumstances: this would allow both host and parasite to survive (Fine 1975; Bull et al. 1991; Wright and Finnegan 2001).

The *Bacillus* stick insect species complex is restricted to the Mediterranean area and shows a number of different reproductive biology issues. The genus comprises three so-called parental species: *Bacillus rossius*, with bisexual and parthenogenetic populations; the strictly bisexual *Bacillus grandii*; and the obligate parthenogenetic *Bacillus atticus*. Interspecific hybridization between or among the parental species produced both unisexual taxa and hybridogenetic lineages (Scali et al. 2003).

To go through the evolution of repeated DNAs, we undertook a genomic sequence survey by randomly cloning genomic fragments of the three parental species *B. rossius*, *B. grandii*, and *B. atticus*, obtaining low coverage DNA libraries. Low coverage sequencing, albeit giving partial genomic information, may provide a quick snapshot of the genome content, especially regarding repetitive DNA. For example, to de novo isolate SINE elements, the random sequencing of a relatively small portion of the genome is a recommended strategy (Nishihara and Okada 2008). In other instances, low coverage genomic surveys (even <0.1×) yield enough data for a good picture of the repeat content (Rasmussen and Noor 2009; Leese et al. 2012).

Here, we present the first results based on repeat masking and de novo characterization of repeated DNAs in genomes of *Bacillus*.

Table 1. Transposable element (TE) families found in the three analyzed genomes.

TE class/ superfamily	TE family	*Bacillus rossius*	*Bacillus grandii*	*Bacillus atticus*
Class I				
LTR		11	5	9
	Bel	2	2	*
	Gypsy	9	3	9
non-LTR		2	2	2
	Jockey	*	1	*
	L2B	n.f.	n.f.	1
	Nimb	1	*	*
	Outcast	*	1	*
	Penelope	1	*	*
	RTE	n.f.	*	1
Class II		16	18	20
	Academ	1	*	*
	Chapaev	*	*	1
	Harbinger	2	*	*
	hAT	5	1	2
	Helitron	1	3	3
	Kolobok	*	2	*
	Mariner	5	6	3
	P	*	1	*
	PiggyBac	*	1	*
	Polinton	2	3	10
	Sola	n.f.	1	1
Total		29	25	31

Note: The number of clones showing significant homology with listed TE families is given. Asterisks indicate TE presence verified through Southern Blot analysis. n.f., not found.

Data presented here will constitute the starting point for further analysis, aiming to clarify the relationships between repetitive DNA sequences and the reproductive biology of the host species.

Materials and methods

Samples and genomic DNA isolation

A *B. rossius* female (Patti, Sicily, gonochoric population), a *B. grandii* male (Ponte Manghisi, Sicily), and a *B. atticus* female (Scoglitti, Sicily) were utilized for the analyses. Gut-deprived specimens were maintained at –80 °C until the DNA isolation was performed through a standard phenol–chloroform procedure.

DNA library construction

For each library, 2 μg of genomic DNA was partially digested with *Eco*RI restriction enzyme (Invitrogen, Carlsbad, Calif., USA) for 2 h at 37 °C. Fragments were ligated to *Eco*RI-adapters (5′-CTCGTAGACTGCGTACC-3′; 5′-AATTGGTACGCAGTCTAC-3′) and then amplified with adaptor-specific primers (5′-GACTGC GTACCAATTCN-3′). After a 1% agarose gel electrophoresis, fragments between 800–1200 bp were recovered by gel extraction and cloned into a pGem-T Easy Vector (Promega, Madison, Wis., USA) used to transform *E. coli* DH5α competent cells (Invitrogen, Carlsbad, Calif., USA). Recombinant colonies were screened by PCR amplification with T7/SP6 primers, under standard PCR conditions. In total, 196 clones per species were sequenced at Macrogen Inc. (Korea).

Southern blot analysis

For each of the 14 TE families missing in at least one of the three libraries, probes were obtained by PCR amplification using specifically designed primers (supplementary data, Table S1)[1]. The PCR program was as follow: initial denaturation at 95 °C for 2 min; 35 cycle of denaturation at 95 °C for 30 s, annealing at 48 °C for 30 s, elongation at 72 °C for 30 s; and a final elongation step at

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/gen-2013-0107.

Table 2. Nucleotide identity of transposable element (TE) families.

TE class	TE family	Species	Avg. identity (%)	*N*	Total matches (bp)
Intraspecific comparisons					
Class I	*Bel*	*Bacillus grandii*	99	2	885
	Gypsy	*Bacillus rossius*	96	7	3340
		Bacillus grandii	97	2	902
		Bacillus atticus	97	3	3063
Class II	*Helitron*	*Bacillus grandii*	66	2	596
		Bacillus atticus	67	2	202
	Mariner	*Bacillus rossius*	99	2	821
		Bacillus grandii	92	5	3630
	Polinton	*Bacillus grandii*	98	2	735
		Bacillus atticus	85	8	3180
Interspecific comparisons					
Class I	*Gypsy*	*Bacillus rossius-grandii*	97	5	4374
		Bacillus rossius-atticus	96	4	2642
		Bacillus grandii-atticus	96	3	1812
Class II	*Helitron*	*Bacillus rossius-grandii*	69	2	326
	Mariner	*Bacillus rossius-atticus*	94	2	867
	Polinton	*Bacillus rossius-atticus*	68	4	797
		Bacillus grandii-atticus	91	6	3054

Note: *N*, number of clones compared.

72 °C for 4 min. PCR reactions were performed with the GoTaq amplification kit (Promega, Madison, Wis., USA), using 30 ng of genomic DNA. Twenty microlitres of each amplification product was separated on a 1.5% agarose gel and Southern blotted onto a positively charged nylon membrane. Hybridization was performed using the AlkPhos labelling and detection kit (GE Healthcare, Pittsburg, Pa., USA) following the manufacturer's protocol. Stringency washes allowed up to 10% of nucleotidic divergence between probes and target DNA.

Sequence analysis

TE identification was done by repeat masking on Repbase Update database with CENSOR web tool (Kohany et al. 2006). We took into account only the hits with nucleotidic score >500 or amino acidic score >300, or those presenting simultaneously amino acidic score >200 and positives >0.5. For family identification, we considered accurate only the alignments with amino acidic score >200 and positives >0.5. Sequence identity for each TE family has been calculated as Σ IA × (LA / LSA), where IA is the identity of the alignment (max identity = 1), LA is the length of the alignment (bp), and LSA is the length of the sum of the alignments (bp).

A de novo search of interspersed repeats was done by self-comparison of each library; an e-value $\leq 10^{-5}$ was set to define significant high scoring segment pairs. Copy number of de novo identified interspersed repeats was calculated following the formula: (No. of occurrences in the library × genome size) / library size.

Differences in relative abundance of scored TEs in the three libraries were tested with repeated measure analysis of variance (ANOVA), followed by post-hoc paired *t* test with Holm correction.

Neighbor-joining tree, using uncorrected *p*-distance, was calculated with MEGA v.5 (Tamura et al. 2011); nodal support was obtained after 500 bootstrap replicates.

Tandem repeat searches were performed by Phobos v.3.3.11 (Mayer 2010), allowing extend exact search, repeat unit size from 2 to 500 bp, and at least four consecutive units.

Sequences were deposited in GenBank under accession numbers KF256266–KF256815.

Results and discussion

For each species, 196 random genomic fragments ~1000 bp long were obtained; on the whole, 144 250, 144 896, and 173 946 bp have been sequenced for *B. rossius*, *B. grandii*, and *B. atticus*, respectively. The genome sizes of these three stick insect species are 2.12–1.90, 2.55–2.11, and 2.25 Gbp, respectively; therefore, the sequencing corresponds to less than 0.006%–0.007% genomic coverage. The average GC content calculated within sequenced libraries is 40.3%, ranging from 39.3% in *B. atticus* to 42.2% in *B. rossius*.

Transposable elements

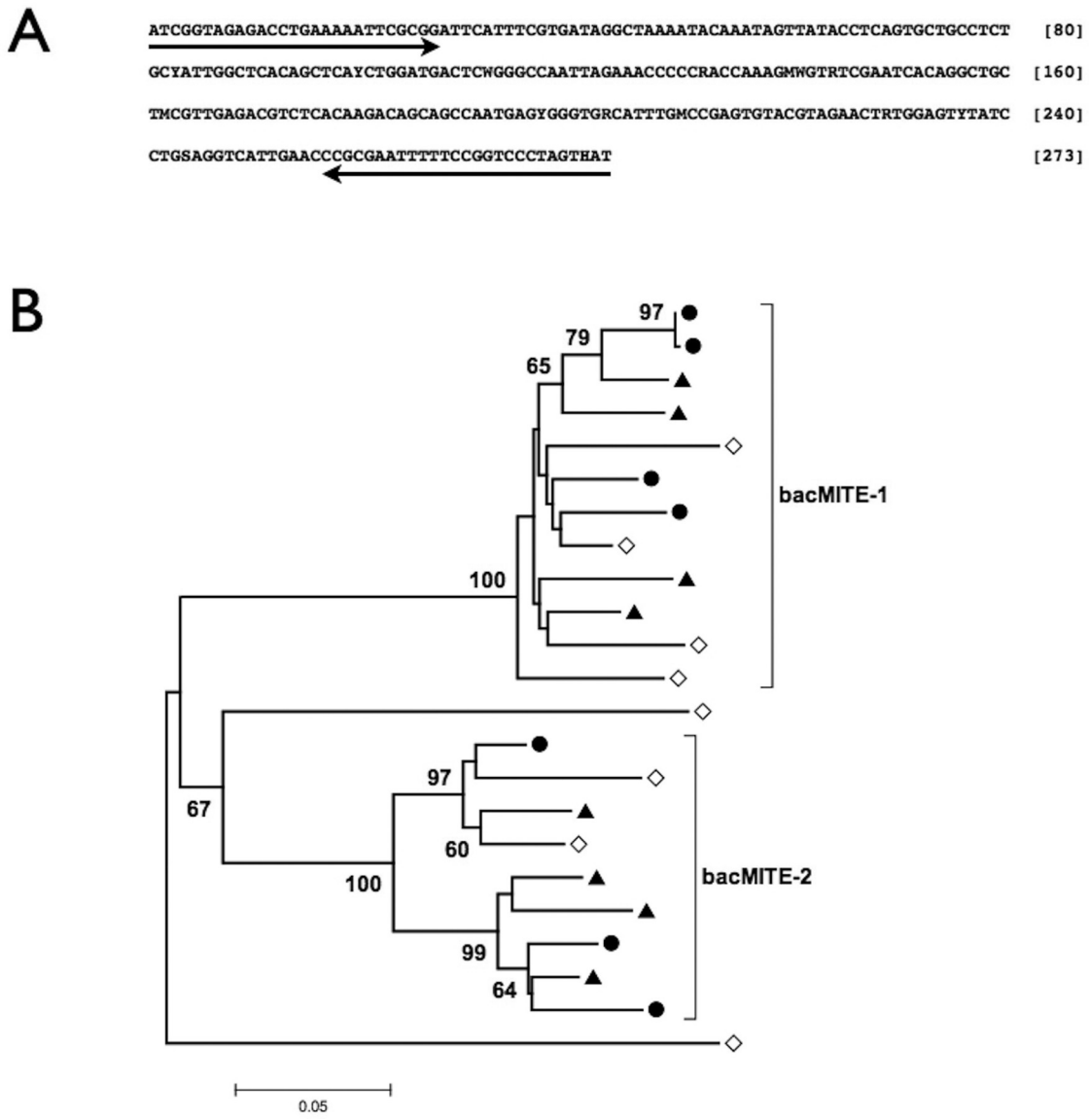
Repeat masking on the three libraries revealed 85 clones containing either class I or class II TEs.

Class I mobile elements belong to the following eight families: *Bel*, *Gypsy*, *Jokey*, *L2B*, *Nimb*, *Outcast*, *Penelope*, and *RTE*. *Bel* and *Gypsy* are LTR retroelements, while the other six are non-LTR retrotransposons (Table 1). Eleven families of class II elements have been identified as follows: *Academ*, *Chapaev*, *Harbinger*, *hAT*, *Helitron*, *Kolobok*, *Mariner*, *P*, *piggyBac*, *Polinton*, and *Sola* (Table 1).

All families have been identified in the three libraries either by clone sequencing or through Southern blot analysis, with the exceptions of *L2B*, *RTE*, and *Sola* in *B. rossius* and *L2B* in *B. grandii* (Table 1). On the whole, the parthenogenetic *B. atticus* shows the presence of all families and shares the majority of them with *B. grandii*, as it could be expected on the basis of phylogenetic relationships (Mantovani et al. 2001). The absence of some families in the gonochoric genomes of *B. rossius* and *B. grandii* may witness the greater ability of bisexuals in dealing with TEs; moreover, considering the time elapsed since the species splitting (23–17 Myr ago; Mantovani et al. 2001), these families could have had enough time for accumulating nucleotide divergence over 10%, and thus becoming undetectable under the Southern blot conditions used in this analysis.

For each library, TE families detected in at least two clones with overlapping regions were analyzed to evaluate the level of similarity (Table 2). These comparisons involved the LTR elements *BEL* and *Gypsy*, and the DNA families *Helitron*, *Mariner*, and *Polinton*. Identity values indicate a substantial intraspecific conservation of sequences, possibly being copies of the same element. The only exception is given by the *Helitron* elements in *B. grandii* and *B. atticus*, with identity values falling to 66% and 67%, respectively. Interspecific comparisons involve the LTR *Gypsy* and the class II elements *Helitron*, *Mariner*, and *Polinton* (Table 2). *Gypsy* and *Mariner* appear well conserved across species, as well as *Polinton* between *B. grandii* and *B. atticus*. On the other hand, a low degree of identity is found for *Helitron* in the *B. rossius* versus *B. grandii* comparison and *Polinton* in the *B. rossius* versus *B. atticus* comparison. On the whole, *Helitron* appears the less conserved element both within

Fig. 1. (A) The 50% majority rule consensus sequences of bacMITE elements. Arrows mark the terminal inverted repeats (TIRs). (B) Neighbor-joining tree based on uncorrected *p*-distance between bacMITE sequences; bootstrap values are calculated after 500 replicates. The two clusters corresponding to the two subfamilies have been indicated. Empty diamonds, *Bacillus rossius* clones; filled circles, *Bacillus grandii* clones; and filled triangles, *Bacillus atticus* clones.



and between genomes: further investigations will clarify if this could be due to higher element diversity.

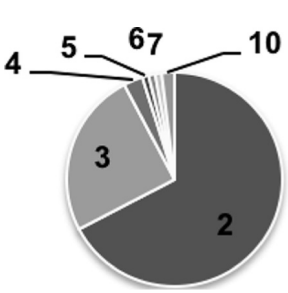
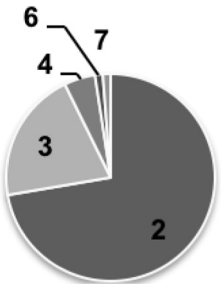
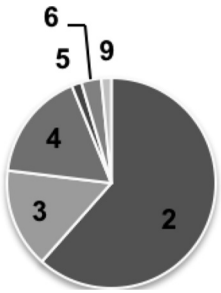
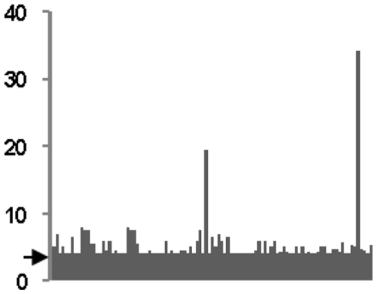
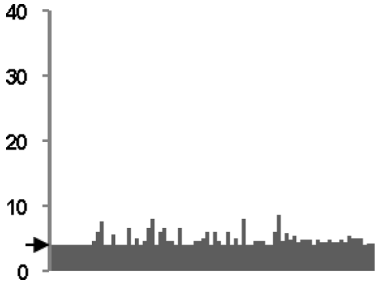
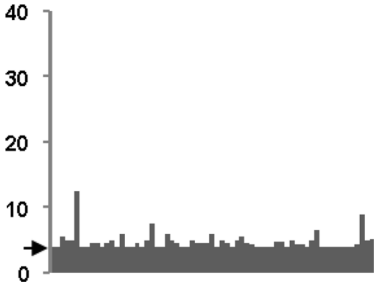
De novo identification of interspersed repeats led to the characterization of 23 homologous nucleotidic stretches 124–277 bp long, distributed in 21 clones from the three genomes, and sharing 77.3% sequence similarity. Aligned sequences gave a consensus length of 273 bp with terminal inverted repeats (TIRs; Fig. 1A). The short length and TIRs are common features of MITEs, small nonautonomous DNA transposon (Feschotte et al. 2002): we, therefore, named the retrieved sequences as bacMITEs. In the phylogenetic analysis, sequences are distributed in two main clusters having 100% nodal support; only two sequences fall outside the two clusters and may represent highly diverging or recombinant elements (Fig. 1B). Therefore, two possible MITE subfamilies can be identified (bacMITE-1 and bacMITE-2), showing within-subfamily sequence identity ranging from 81.5% to 91.2%. The relative copy numbers are 5.6×10^4, 4.8×10^4, 5.6×10^4, for bacMITE-1, and 5.6×10^4, 6.4×10^4, 5.6×10^4, for bacMITE–2, in *B. rossius*, *B. grandii*, and *B. atticus*, respectively. No flanking target site duplications (TSD) have been identified, but this could be due to the low number of identified full-length elements (three bacMITE-1 and one bacMITE-2). To our knowledge, bacMITEs are the first MITE elements ever found within a polyneopteran genome.

It is worth noting that no SINEs have been identified in this survey. Of course, this could be due to the limited sequencing

Table 3. Occurrences of transposable element (TE), relative percentages of TE classes among positive clones, and class II/class I ratio.

	Bacillus rossius	*Bacillus grandii*	*Bacillus atticus*
Total TEs/library	23.3%	18.0%	22.9%
Class I			
LTR	27.5%	11.8%	20.5%
non-LTR	10.0%	14.7%	6.8%
Class II	62.5%	73.5%	72.7%
Class II/class I	1.667	2.778	2.667

Table 4. Tandem repeat abundance (bp) detected in the three libraries.

	Bacillus rossius	*Bacillus grandii*	*Bacillus atticus*
Microsatellite	1.00% (103)	0.64% (83)	0.54% (65)
Unit length distribution (bp)			
Array length distribution (bp)			
Minisatellite	0.32% (2)	0.00% (0)	0.36% (3)
Unit length/copy number	15/11 45/7 —	n.f. — —	16/8 19/6 30/10
Satellite	0.00% (0)	0.00% (0)	0.96% (2)
Total	1.32% (105)	0.64% (23)	1.86% (70)

Note: The number of loci is given in parentheses. For microsatellite loci, array distributions are also shown (*x* axis, unit length; *y* axis, copy number; arrow indicates the four copy threshold used as the minimum tandem array length during Phobos v.3.3.11 search; each bin represents a single locus). n.f., not found.

relative to the genome size of *Bacillus* spp., but it should be noted that, in other instances, low coverage sequencing gave at least some SINE sequences. For example, in the termite *Reticulitermes lucifugus*, 25 SINE sequences belonging to four distinct SINE families have been found by random sequencing ~130 000 bp (approximately 0.012% of the termite genome; Luchetti and Mantovani 2011). In the hyrax genome, the random sequencing of 63 000 bp (~0.002% of the whole genome) revealed 26 AfroSINE elements (Nishihara and Okada, 2008). It is, thus, possible to hypothesize that the genome of *Bacillus* lacks SINEs, as observed in species of *Drosophila* (Kramerov and Vassetzky 2011), or that they are very poorly represented. Further sequencing or focused experiments on SINE search will allow more definitive conclusions.

As a general picture, LTR, non-LTR, and DNA transposons have quite different representativeness ($p = 0.002$), though this abundance distribution is not significantly different among the surveyed libraries ($p = 0.521$).

DNA transposons are generally prevailing in all libraries (class II versus LTR, $p = 0.00060$; class II versus non-LTR, $p = 0.00032$), their amount ranging from 1.666- to 2.778-fold higher than that of retrotransposons. This variability is not unexpected, taking into account that the class II/class I ratio may vary from ~100% class II to ~100% class I elements. For example, in two closely related insect species, the Culicidae mosquitoes *Anopheles gambiae* and *Aedes aegypti*, this ratio ranges from ~0.7 to ~2.3, respectively (reviewed in Feschotte and Pritham 2007). Within class I elements, LTR families outnumber non-LTR families (Table 3), but this difference is not significant ($p = 0.17433$).

Tandem repeats

On the whole, less than 2% of the sequenced genomic fragments in each species of *Bacillus* is constituted by tandem repeats.

As expected, microsatellite loci (2–10 bp) are more represented than minisatellite or satellite DNA. Di- and trinucleotide loci are the most abundant, with few instances of longer units (Table 4). However, most of the retrieved loci retains the minimum length imposed during the repeat search or remains below the 10 copies; only two loci in *B. rossius* and one locus in *B. atticus* showed longer arrays. In the former species, a dinucleotide array showed 19 repeats and a pentanucleotide locus has up to 34 repeat units. In the latter taxon, a dinucleotide array is made by 12 repeats (Table 4).

Minisatellites (11–100 bp) occurred only in *B. rossius* and *B. atticus* libraries, with repeat unit length ranging from 16 to 45 bp and copy number comprised between 6 and 10 (Table 4). Interestingly, a 45-mer array found in *B. rossius* has significant homologies with clones from both *B. grandii* and *B. atticus* libraries. Sequence analysis indicates that one clone from *B. grandii* (gra_af4) and five from *B. atticus* (att_ab7, att_ah5, att_ah7, att_ah11, and att_bc8) have from two to three tandemly arranged 45-mer repeat units, with an overall repeat units sequence identity of 62%. Therefore, the 5′ and the 3′ array flanking regions were compared to check if the same genomic locus harbours this minisatellite in all the three genomes: while the 3′ end flanking region did not show any homology among clones, the 5′ flanking region was significantly conserved even between species (74.1% pairwise identity; 84.4% of identical sites), with the exception of the clone att_ah11. The analysis of the consensus sequence generated from the alignment of 5′ flanking regions did not give any similarity with any known sequence in public databases. Tandem repeat arrays flanked by

homologous sequences are, actually, commonly found: in fact, they can originate within the TE sequence (Mogil et al. 2012; Sharma et al. 2013) or, in most cases, they are generated at retrotransposon tails when they reintegrate into new genomic locations (Lopez-Giraldez et al. 2006; Meglécz et al. 2007; Coates et al. 2009, 2011; Luchetti and Mantovani 2009, 2011). However, retrotransposons usually generate microsatellite loci, while in this case a 45-bp unit made the tandem array. At present, it is not possible to further explain such occurrence, especially because the homologous flanking sequence is not similar to any known TE sequence. However, it would be interesting to check whether this minisatellite locus has been generated upon the insertion of a retrotransposon or by some recombinative mechanism during the evolution of the genome in species of the genus *Bacillus*.

Satellite DNA sequences (>100 bp) have been retrieved only in the *B. atticus* library, where five *Bag320* monomers (Mantovani et al. 1997; Cesari et al. 2003; Luchetti et al. 2003) have been found in two clones (att_ac9 and att_ba1). With respect to previously isolated *Bag320* sequences, they show a sequence similarity with *B. atticus* specific monomers ranging from 94.1% to 96.0%. No *Bag320* sequences have been found in the genomes of *B. rossius* or *B. grandii*. For the former species, this was quite expected, as it is known that this satellite DNA occurs at a very low copy number in this genome and was only isolated by PCR amplification (Cesari et al. 2003). Its absence in the *B. grandii* library is, however, more surprising because this genome contains the highest copy number of the satellite family (15%–20% in *B. grandii* versus 2%–5% in *B. atticus*; Mantovani et al. 1997). As a general consideration, the use of *Eco*RI restriction enzyme for library production may have biased the genomic sampling of *Bag320* sequences, as they do not contain its cutting site in the considered species; therefore, its sampling from *B. atticus* genomes could be considered as mere chance.

On the whole, the sequenced libraries represent a small fraction of the whole genome of *B. rossius*, *B. grandii*, and *B. atticus*, with less than 1% genomic coverage. Yet, the survey allowed the retrieval of a number of repetitive DNAs, either interspersed or not. Most of the main TE families are represented, and a MITE family, the first ever discovered in polyneopteran insects, has been de novo characterized. Moreover, mini- and microsatellite loci were found even if characterized by short arrays. Unfortunately, it is not possible to make strong comparisons on the representativeness of repetitive DNAs within the sequenced libraries, as no polyneopteran genomes have been sequenced so far and the only polyneopteran DNA library available has been built from the termite species *R. lucifugus*. In this library (covering 0.012% of the genome), four SINE and one putative MITE families were de novo characterized (Luchetti and Mantovani 2011; A. Luchetti, unpublished data); moreover, 11 minisatellite and 298 microsatellite loci were isolated in a single termite species. Therefore, despite the smaller genome (~1 versus >2 Gbp), termites appear to have a more repetitive genomic landscape than stick insects. As a final remark, it is interesting to point out the relative TE content showed by unisexual (*B. atticus*) as opposed to gonochoric taxa (*B. rossius* and *B. grandii*). Both theoretical and empirical studies evidenced that parthenogenetic, i.e., low recombining, genomes should avoid TE accumulation to escape the effects of Muller's ratchet; otherwise, TE load would raise without the possibility of elimination eventually leading to the host species extinction (Arkhipova and Meselson 2000, 2005; Sullender and Crease 2001; Dolgin and Charlesworth 2006). On the other hand, in the parasitoid wasp *Leptopilina clavipes*, unisexual and bisexual lineages showed no difference in the overall TE genomic coverage (Kraaijeveld et al 2012). In line with this, *B. atticus* genome appears to have (at least) the same TE content of the bisexual species. As argued by Wright and Finnegan (2001), TE prevalence studies in obligate unisexual genomes are difficult to interpret mainly because they are derived from bisexual ancestors. Thus, at the moment, it is impossible to state whether *B. atticus* and *L. clavipes* TEs have been selected for less harmful elements or they are still in the "elimination phase", i.e., the two taxa are not unisexual since enough time to allow an efficient clearance of TE load. On the whole, although based on small datasets and considering the different reproductive strategies, the variation of ploidy, and the presence of hybrid taxa that characterize the *Bacillus* complex, the present survey provides interesting preliminary data to undertake further analysis in species of the genus *Bacillus*.

Acknowledgements

This work was funded by Canziani and RFO funds to B.M. M.R. produced the three libraries and carried out repeat masking; A.L. performed de novo TE searches and tandem repeat characterization; L.B. performed Southern blot analysis; B.M. supervised the project; and M.R., A.L., and B.M. wrote the article. Authors wish to thank the two reviewers whose criticisms and suggestions helped to substantially improve the paper.

References

Arkhipova, I., and Meselson, M. 2000. Transposable elements in sexual and ancient asexual taxa. Proc. Natl. Acad. Sci. U.S.A. **97**(26): 14473–14477. doi:10.1073/pnas.97.26.14473. PMID:11121049.

Arkhipova, I., and Meselson, M. 2005. Deleterious transposable elements and the extinction of asexuals. Bioessays, **27**(1): 76–85. doi:10.1002/bies.20159. PMID:15612027.

Brookfield, J.F.Y., and Badge, R.M. 1997. Population genetics models of transposable elements. Genetica, **100**(1–3): 281–294. doi:10.1023/A:1018310418744. PMID:9440281.

Bull, J.J., Molineux, I.J., and Rice, W.R. 1991. Selection of benevolence in a host-parasite system. Evolution, **45**(4): 875–882. doi:10.2307/2409695.

Cesari, M., Luchetti, A., Passamonti, M., Scali, V., and Mantovani, B. 2003. Polymerase chain reaction amplification of the *Bag320* satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). Gene, **312**: 289–295. doi:10.1016/S0378-1119(03)00625-5. PMID:12909366.

Charlesworth, B., Sniegowski, P., and Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature, **371**(6494): 215–220. doi:10.1038/371215a0. PMID:8078581.

Coates, B.S., Sumerford, D.V., Hellmich, R.L., and Lewis, L.C. 2009. Repetitive genome elements in a European corn borer, *Ostrinia nubilalis*, bacterial artificial chromosome library were indicated by bacterial artificial chromosome end sequencing and development of sequence tag site markers: implications for lepidopteran genomic research. Genome, **52**(1): 57–67. doi:10.1139/G08-104. PMID:19132072.

Coates, B.S., Kroemer, J.A., Sumerford, D.V., and Hellmich, R.L. 2011. A novel class of miniature inverted repeat transposable elements (MITEs) that contain hitchhiking $(GTCY)_n$ microsatellites. Insect. Mol. Biol. **20**(1): 15–27. doi:10.1111/j.1365-2583.2010.01046.x. PMID:20977507.

Dolgin, E.S., and Charlesworth, B. 2006. The fate of transposable elements in asexual populations. Genetics, **174**(2): 817–827. doi:10.1534/genetics.106.060434. PMID:16888330.

El-Murr, N., Abidi, Z., Wanherdrick, K., Svrcek, M., Gaub, M.P., Fléjou, J.F., et al. 2012. *MiRNA* genes constitute new targets for microsatellite instability in colorectal cancer. PLoS ONE, **7**(2): e31862. doi:10.1371/journal.pone.0031862. PMID:22348132.

Feschotte, C., and Pritham, E. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. **41**: 331–368. doi:10.1146/annurev.genet.40.110405.090448. PMID:18076328.

Feschotte, C., Zhang, X., and Wessler, S.R. 2002. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. *In* Mobile DNA II. *Edited by* N. Craig, R. Craigie, M. Gellert, and A. Lambowitz. A.S.M. Press, Washington, D.C. pp. 1147–1158.

Fine, P.E.M. 1975. Vectors and vertical transmission: an epidemiologic perspective. Ann. N.Y. Acad. Sci. **266**: 173–194. doi:10.1111/j.1749-6632.1975.tb35099.x. PMID:829470.

Galindo, C.L., McIver, L.J., Tae, H., McCormick, J.F., Skinner, M.A., Hoeschele, I., et al. 2011. Sporadic breast cancer patients' germline DNA exhibit an AT rich microsatellite signature genes chromosomes. Cancer, **50**(4): 275–283. PMID:21319262.

Kazazian, H.H. 2004. Mobile elements: drivers of genome evolution. Science, **303**(5664): 1626–1632. doi:10.1126/science.1089670. PMID:15016989.

Kohany, O., Gentles, A.J., Hankus, L., and Jurka, J. 2006. Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. BMC Bioinformatics, **25**(7): 474. PMID:17064419.

Kraaijeveld, K., Zwanenburg, B., Hubert, B., Vieira, C., De Pater, S., Van Alphen, J.J.M., et al. 2012. Transposon proliferation in an asexual parasi-

toid. Mol. Ecol. **21**(16): 3898–3906. doi:10.1111/j.1365-294X.2012.5582.x. PMID: 22548357.

Kramerov, D.A., and Vassetzky, N.S. 2011. Origin and evolution of SINEs in eukaryotic genomes. Heredity, **107**(6): 487–495. doi:10.1038/hdy.2011.43. PMID: 21673742.

Leese, F., Brand, P., Rozenberg, A., Mayer, C., Agrawal, S., Dambach, J., et al. 2012. Exploring Pandora's box: potential and pitfalls of low coverage genome surveys for evolutionary biology. PLoS ONE, **7**(11): e49202. doi:10.1371/journal.pone.0049202. PMID:23185309.

López-Giráldez, F., Andrés, O., Domingo-Roura, X., and Bosch, M. 2006. Analyses of carnivore microsatellites and their intimate association with tRNA-derived SINEs. BMC Genomics, **7**: 269. doi:10.1186/1471-2164-7-269.

Luchetti, A., and Mantovani, B. 2009. *Talua* SINE biology in the genome of the *Reticulitermes* subterranean termites (Isoptera, Rhinotermitidae). J. Mol. Evol. **69**(6): 589–600. doi:10.1007/s00239-009-9285-7. PMID:19904483.

Luchetti, A., and Mantovani, B. 2011. Molecular characterization, genomic distribution and evolutionary dynamics of Short INterspersed Elements in the termite genome. Mol. Genet. Genomics, **285**(2): 175–184. doi:10.1007/s00438-010-0595-7.

Luchetti, A., Cesari, M., Carrara, G., Cavicchi, S., Passamonti, M., Scali, V., and Mantovani, B. 2003. Unisexuality and molecular drive: *Bag320* sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). J. Mol. Evol. **56**(5): 587–596. doi:10.1007/s00239-002-2427-9. PMID:12698295.

Makałowski, W., Pande, A., Gotea, V., and Makałowska, I. 2012. Transposable elements and their identification. Methods Mol. Biol. **855**: 337–359. doi:10.1007/978-1-61779-582-4_12. PMID:22407715.

Mantovani, B., Tinti, F., Bachmann, L., and Scali, V. 1997. The *Bag320* satellite DNA family in *Bacillus* stick insects (Phasmatodea): different rates of molecular evolution of highly repetitive DNA in bisexual and parthenogenetic taxa. Mol. Biol. Evol. **14**(12): 1197–1205. doi:10.1093/oxfordjournals.molbev.a025729. PMID:9402731.

Mantovani, B., Passamonti, M., and Scali, V. 2001. The mitochondrial cytochrome oxidase II gene in *Bacillus* stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. Mol. Phylogenet. Evol. **19**(1): 157–163. doi:10.1006/mpev.2000.0850. PMID:11286500.

Mayer, C. 2010. Phobos 3.3.11. Available from http://worldwidewords.rub.de/spezoo/cm/cm_phobos.htm [accessed 22 April 2013].

Meglécz, E., Anderson, S.J., Bourguet, D., Butcher, R., Caldas, A., Cassel-Lundhagen, A., et al. 2007. Microsatellite flanking region similarities among different loci within insect species. Insect. Mol. Biol. **16**(2): 175–185. doi:10.1111/j.1365-2583.2006.00713.x. PMID:17298557.

Mogil, L.S., Slowikowski, K., and Laten, H.M. 2012. Computational and experimental analyses of retrotransposon-associated minisatellite DNAs in the soybean genome. BMC Bioinformatics, **13**(Suppl. 2): S13. doi:10.1186/1471-2105-13-S2-S13. PMID:22536864.

Nishihara, H., and Okada, N. 2008. Retroposons: genetic footprints on the evolutionary paths of life. Methods Mol. Biol. **422**: 201–225. doi:10.1007/978-1-59745-581-7_13. PMID:18629669.

Plohl, M., Luchetti, A., Meštrović, N., and Mantovani, B. 2008. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene, **409**(1–2): 72–82. doi:10.1016/j.gene.2007.11.013. PMID:18182173.

Rasmussen, D.A., and Noor, M.A.F. 2009. What can you do with 0.1× genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). BMC Genomics, **10**: 382. doi:10.1186/1471-2164-10-382.

Richard, G.-F., Kerrest, A., and Dujon, B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol. Mol. Biol. Rev. **72**(4): 686–727. doi:10.1128/MMBR.00011-08. PMID:19052325.

Scali, V., Passamonti, M., Marescalchi, O., and Mantovani, B. 2003. Linkage between sexual and asexual lineages: genome evolution in *Bacillus* stick insects. Biol. J. Linn. Soc. **79**: 137–150. doi:10.1046/j.1095-8312.2003.00172.x.

Sharma, A., Wolfgruber, T.K., and Presting, G.G. 2013. Tandem repeats derived from centromeric retrotransposons. BMC Genomics, **14**: 142. doi:10.1186/1471-2164-14-142. PMID:23452340.

Sullender, B.W., and Crease, T.J. 2001. The behavior of a *Daphnia pulex* transposable element in cyclically and obligately parthenogenetic populations. J. Mol. Evol. **53**(1): 63–69. PMID:11683324.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol. Biol. Evol. **28**: 2731–2739. doi:10.1093/molbev/msr121. PMID:21546353.

Wright, S., and Finnegan, D. 2001. Genome evolution: sex and the transposable element. Curr. Biol. **11**: R296–R299. doi:10.1016/S0960-9822(01)00168-3. PMID: 11369217.

Wright, S.I., and Schoen, D.J. 1999. Transposon dynamics and the breeding system. Genetica, **107**(1–3): 139–148. doi:10.1023/A:1003953126700. PMID: 10952207.