Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BIODIVERSITA' ED EVOLUZIONE

Ciclo XXVIII

Settore Concorsuale di afferenza: 05/B1 – Zoologia ed Antropologia

Settore Scientifico disciplinare: BIO/05 – Zoologia

TRANSPOSABLE ELEMENTS IN ARTHROPODS GENOMES WITH NON-CANONICAL REPRODUCTIVE STRATEGIES.

Presentata da: Dot.ssa Claudia Scavariello

Coordinatore Dottorato

Prof.ssa Barbara Mantovani

Relatore

Prof.ssa Barbara Mantovani

Correlatore

Dott. Andrea Luchetti

Esame finale anno 2016

INDEX

1. INTRODUCTION

1.1. History of transposable elements

Until the half of the 20th century, the genome was basically considered like a pearl necklace with genes ordered on chromosomes, behaving as stable and almost fixed entities evolving through the accumulation of random mutations at slow rate. Only in the late ’40s, Barbara McClintock was able to demonstrate that genomes are highly dynamic entities thanks to her discovery that some genes could be able to move from a chromosomal site to another one. Genetic studies on maize led McClintock to characterize a DNA sequence that could move from one chromosome location to another causing chromosomal breakage or instability, and leading to chromosomal rearrangements. She also found that in one maize line the breakage on the short arm of chromosome 9 was recurrent. She assumed that these events were from the product of a particular unstable genetic element named *Ds* (i.e., Dissociation). Through genetic crosses and cytological observations, she understood that the instability of *Ds* elements was dependent on the presence of another element type designated as *Ac* for Activator.

Her data on the first transposable elements (TEs) ever found, *Ac* and *Ds*, were reported in “The origin and behavior of mutable loci in maize”, article published in 1950 (McClintock, 1950). In the 1980s, it was understood that *Ac* and *Ds* are autonomous and non-autonomous elements and only *Ac* encodes the functional transposase enzyme required for the mobility of both elements (McClintock, 1953).

Only twenty years later McClintock's discoveries, the scientific community started to acknowledge her assumptions, thanks to the work of the American microbiologist James A. Shapiro, the American geneticist Margaret Kidwell and French biologist Georges Picard. Shapiro worked on the galactose (*gal*) operon in *E. coli.* He isolated a *gal*-mutants in the late 1960s and he demonstrated that the mutations were caused by the recurrent insertions of the same long segment of DNA, called insertion sequence 1 (IS1). All IS1 were very similar in structure: they had a single gene sequence encoding a transposase enzyme and delimited by short terminal inverted repeats (TIRs) (Shapiro 1969). Kidwell and Picard discovered the presence of P elements in the *Drosophila melanogaster* genome through studies on hybrid dysgenesis. In the hybrids of crosses between females from laboratory stocks with males from specific natural populations (called P strains), dysgenesis induced frequent sterility and genetic instability, but curiously the same situation was not observed in the reciprocal crosses (Kidwell 1977). The molecular characterization of mutations in these hybrids established that hybrid dysgenesis was

caused by a mobile element activated during dysgenic crosses: this P element was consistently absent from laboratory stocks but present in most wild-caught flies (Picard et al. 1978; Kidwell 1979). Like IS elements, P elements have a short TIRs and a single gene encoding a transposase.

In the early 1980s, studies on another popular model organism, the yeast *Saccharomyces cerevisiae,* revealed the presence of TEs, the so-called Ty elements, completely different from other mobile elements known at that time. These elements showed long terminal repeats (LTRs) flanking an array of coding sequences and some common characteristics with retroviral proviruses, such as a reverse transcriptase domain. An American geneticist, Jef Boeke, confirmed that Ty elements transpose through an RNA intermediate which is converted back into a DNA molecule for its integration in the yeast genome. Boeke coined the term retrotransposon to describe TEs that move via this process (Boeke et al. 1985).

In the 1980s, further, appropriate molecular tools were developed for eukaryotic systems and many non-model organisms started to be investigated, highlighting that TEs were universal components of all living organisms.

During those years, numerous empirical data accumulated about TEs structure and transposition mechanisms but little had been speculated on theoretical expectations.

At that time, TEs studies were influenced by two different theories trying to explain the ubiquitous presence of TEs as well as their high genomic proportions. The first theory that influenced TEs studies was the neutral theory at the end of the 1970s and early 1980s, of whom Dawkins and Crick were the main leaders. This theory considered these elements as "*junk*'' and ''*selfish*'' pieces of DNA whose only "function" was self-preservation (Dawkins, 1976; Crick, 1979; Doolittle and Sapienza, 1980). In his book *The Selfish Gene* Richard Dawkins (1976) suggested the idea that surplus non-coding DNA in eukaryotic genomes was selfish. Selfish DNA is a DNA sequence with the unique apparent function to replicates itself, also exploiting the host molecular mechanisms, this is why it was considered a 'molecular parasite' in the selfish DNA theory. The presence of selfish DNA can be explained by the inability of the organism to efficiently prevent its increase in the genome.

Another theory explained instead the large presence of TEs in living organisms as due to some advantage they perform for their host genomes. The "phenotypic paradigm" of the neo-Darwinian theory speculated that non-transcribed DNA, and in particular repetitive sequences, were necessary as regulators or somehow essential to chromosome structures or pairing. Under this view, this genomic portion would facilitate genetic rearrangements increasing the possibility of genome evolution (long-term phenotypic

benefit); or it could evolve into sequences with new functions (McDonald, 1983).

A very interesting evolutionary aspect of TE biology started to emerge in the early '80s when Temin (1980) and Flavell (1981) highlighted a possible link between retroviruses and retroelements. In fact, they realized that LTR retrotransposons, in particular, share some common structures with retroviruses, such as LTR, *gag*, and *pol* genes and, in some cases, incomplete and nonfunctional *env* genes. Since then, a number of fundamental evolutionary studies started. Studies based on reverse transcriptase similarities (Xiong and Eickbush,1990; McClure, 1991) and integrase/transposase comparison (Capy et al.,1996) suggested that retroviruses may have evolved from LTR retrotransposons by acquiring a functional *env* gene. However, comparative studies on gypsy elements in *Drosophila melanogaster* and *Drosophila subobscura* demonstrated that the gypsy ORF3 of *D. subobscura*, while presenting the whole structure, encodes a non-functional *env* protein. This result launched the hypothesis that some gypsy elements may have lost functional *env* gene and consequently they lost infective ability. In this way, an opposite view – i.e. *D. subobscura* gypsy elements are degenerate forms of insect retroviruses – was proposed (Alberola et al., 1997). The debate over retroviruses/LTR retrotransposons origin is still ongoing.

The initial findings on TEs replication mechanisms led to their first classification mainly on the way they transpose. The elements transposing via a DNA intermediate, were named ‘‘transposons’’, while the elements transposing through an RNA intermediate, were called “retrotransposons”. The retrotransposons with long terminal repeats at their ends were called LTR retrotransposons to distinguish them from retrotransposons lacking these long terminal repeats, the non-LTR retrotransposons.

This simple classification has been used for many years. Only in the early 21th century Wicker et al. (2007) and Kapitonov and Jurka (2008) proposed a new TE classification based not only on transposition mechanism, but also on sequence similarities and reverse transcriptase phylogenetic relationships to classify more easily all the new emerging TEs classes.

1.1.1 TEs Classification

TEs classification is something far from being settled and poses many problems also owing to the increasing number of available genomic sequences (Piégu et al., 2015). From a very general point of view a first distinction can be made on the intermediates of their transposition process (Fig. 1). Class I comprises retrotransposons; they move through RNA intermediates, following a process analogous to a “copy-and-paste” mechanism.

These elements, in fact, code for a reverse transcriptase (RT) which synthesizes DNA starting from the transcribed mRNA; the DNA copies produced are then re-integrate into new sites of the genome. Class II includes transposons: they move by DNA intermediates, this mode of transposition being indicated as a “cut-and-paste” mechanism. Transposition starts through a transposase which binds inverted repeats located at the transposon ends. In this way, the enzyme produces a stable nucleoprotein complex, said transpososome. The transpososome ensures that DNA cutting and assembly reactions, necessary for transposon movement, are carried out simultaneously at the two ends of the DNA element. Three types of eukaryotic class I TEs can be further distinguished: Long Terminal repeat (LTR) elements (DNA integration occurs with a transposase-related integrase); non-LTR elements (DNA integration and RT priming is mediated by an endonuclease) and *Dictyostelium* Intermediate Repeat Sequence 1 (DIRS-1) – like elements (DNA integration uses a tyrosine recombinase; Choen et al., 1984; Chenais et al., 2012). LTR retrotransposons have long direct terminal repeats and from 1 to 3 open reading frames (ORFs) similar to the *gag* and *pol* retroviruses genes; the *gag* gene produces a protein for DNA binding, while the *pol* gene has several protein domains including RT domain. The non-LTR retrotransposons lack LTRs, but typically have a poly(A) tail at the 3' end and deletions of variable length at the 5' end. They include autonomous Long INterspersed Elements (LINEs) and Short INterspersed Elements (SINEs), the latter being non autonomous elements. LINEs are divided into two classes, based on their structure. The first class is characterized by two ORF. The first ORF produces a protein for RNA binding, while the second ORF contains the RT domain and the endonuclease domain (Wessler, 2006). The second class has a single ORF in which the RT domain is centrally located. The RT is flanked by a N-terminal domain with DNA binding motif(s) (eg. Zinc Finger, ZF) and a C-terminal domain for endonuclease activities.

SINEs, as all non-autonomous elements, must use the enzymes synthesized by autonomous elements for their transposition (Kapitonov and Jurka, 2008). SINEs in particular depend on partner LINEs enzymatic machinery in order to achieve reverse transcription and integration. In fact, some SINEs share the same sequence at their 3′ end regions with the partner LINE found in the same genome (Jurka 1997).

The majority of DIRS-like elements have a structure similar to LTR retrotransposons but, unlike the latter, they have a different insertion mechanism into the target site: they use a recombinase rather than an integrase. Also, if present, they have inverse long terminal repeats.

DNA transposons (class II) use, for their transposition, a DNA intermediate which can be

either single or double stranded. These transposons can be divided into three major types: type 1 elements have two terminal inverted repeats (TIRs) and are typical “cut-and-paste” DNA transposons, as described above, with a double stranded DNA intermediate and DNA transposase as leading enzyme. Type 2 elements are rolling-circle DNA transposons also known as Helitrons (Kapitonov and Jurka, 2007): the transposition of these elements follows a semi-replicative transposition model, because only one strand of the element is transferred from its site to new target site and it is used as a template for DNA synthesis catalyzed by the host repair machinery.

Type 3 elements, also known as Polintons or Mavericks, are self-synthesizing DNA transposons, whose transposition is coupled with their own DNA synthesis, catalyzed by the Polinton-encoded DNA polymerase (Kapitonov and Jurka, 2006). Excision and integration of Polintons are catalyzed by “cut and paste”-like DNA transposases similar to retroviral integrases. Both Helitrons and Polintons and Mavericks produce a single strand cut.

Class II non-autonomous elements are, mainly, Miniature Inverted repeat Transposable Element (MITE). They derive from autonomous elements through deletion of the internal protein-coding sequence, have a small size (50-500 bp long) but maintain the presence of terminal inverted repeats (TIRs). Initially, they were characterized in plants, but they are also present in metazoan (Feschotte et al. 2002).

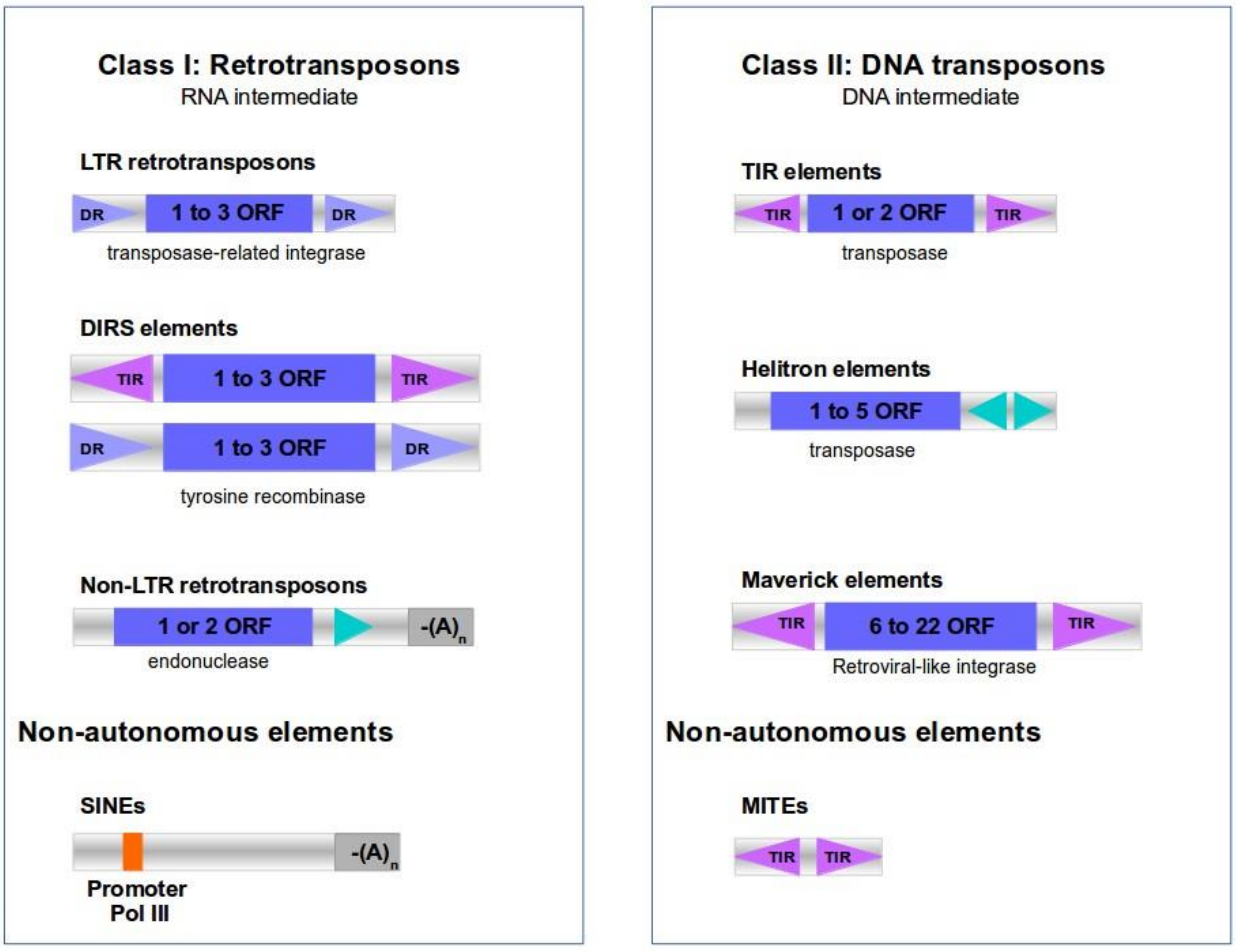


Fig.1 TEs classification. The triangles indicate the DNA repeats. The number of ORF is shown.

1.2. R2 non-LTR retrotransposon

1.2.1. R2 history and structure

In the late 1970s, genetic studies in *Drosophila melanogaster* had shown that a large proportion of 28S ribosomal genes contained insertions in their sequence (Wellauer and Dawid, 1977). These insertions had repressive effects on rDNA transcription. On the basis of their junction regions they were divided into two classes. Type I elements are inserted at a position around two-thirds of the distance from the 5' end of the gene and are flanked by a 14 base pair (bp) duplication. Type II elements interrupt the 28S gene 75 bp upstream of the type I elements and do not present flanking duplications (Roiha and Glover, 1981). In 1987 Burke et al. (1987) suggested to change the element names to avoid a possible confusion with intron sequences of fungal mitochondrial DNA also called type I and type II elements. Since these insertions are within ribosomal (R) DNA, they were called R1 and R2, respectively.

R elements are non-LTR retrotransposons and nine R retrotransposon families have been so far identified (Fig. 2): R1, R2, R4, R5, R6, R9, RT are inserted into the 28S rDNA locus,

while R7 and R8 interrupt the 18S sequence (Eickbush e Eickbush, 2007). Structural and phylogenetic analyses demonstrated that R1, R6, R7, R9 and RT encode an endonuclease, upstream the reverse transcriptase sequence, related to apurinic endonucleases; they are all classified into the R1 clade (Kojima and Fujiwara, 2003).

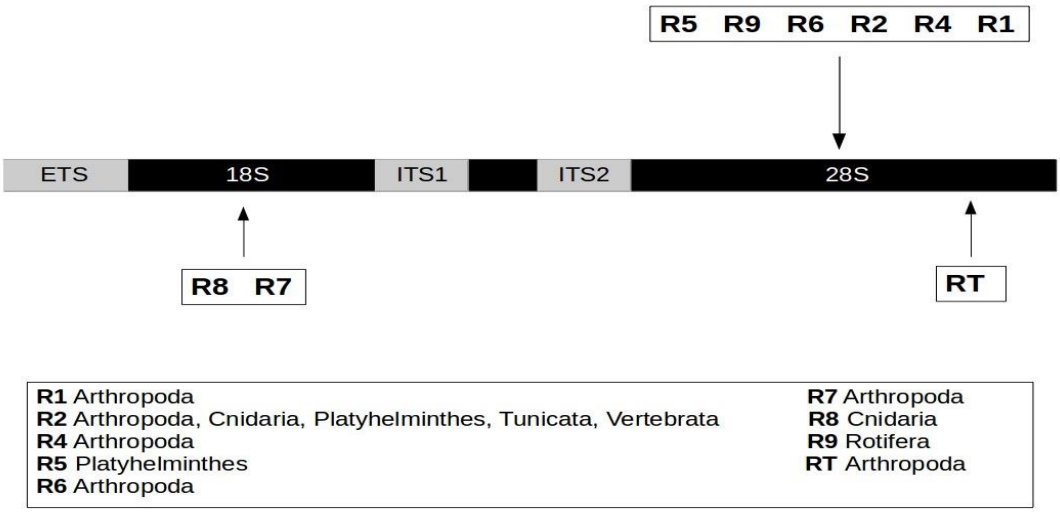


Fig. 2. R elements series. Their insertion site and hosts are indicated.

R2, R4, R5, and R8 encode a restriction enzyme-like endonuclease (RLE) downstream the reverse transcriptase domain. These retroelements appear to descend from a common lineage found in many animal taxa (Kojima et al., 2006).

At present, R2 is one of the most site-specific non-LTR retrotransposon family studied. It specifically inserts into the sequence 5'-TTAAGG↓TAGCCA-3' of the 28S rRNA gene. R2 structure (Fig. 3) comprises a single ORF flanked by two untranslated regions (UTR). The ORF encodes a N-terminal DNA-binding domain, a central RT domain and a endonuclease domain at the C-terminus (Christensen et al., 2006). The N-terminal domain can contain one (CCHH), two (CCHH + CCHH or CCHC + CCHH) or three (CCHH + CCHC + CCHH) cysteine-histidine motifs (zinc finger) and in addition there is a Myb motif, involved in the DNA binding, while the C-terminal domain has only one cysteine-histidine motif (CCHC) (Kojima and Fujiwara, 2005; Luchetti and Mantovani, 2013).

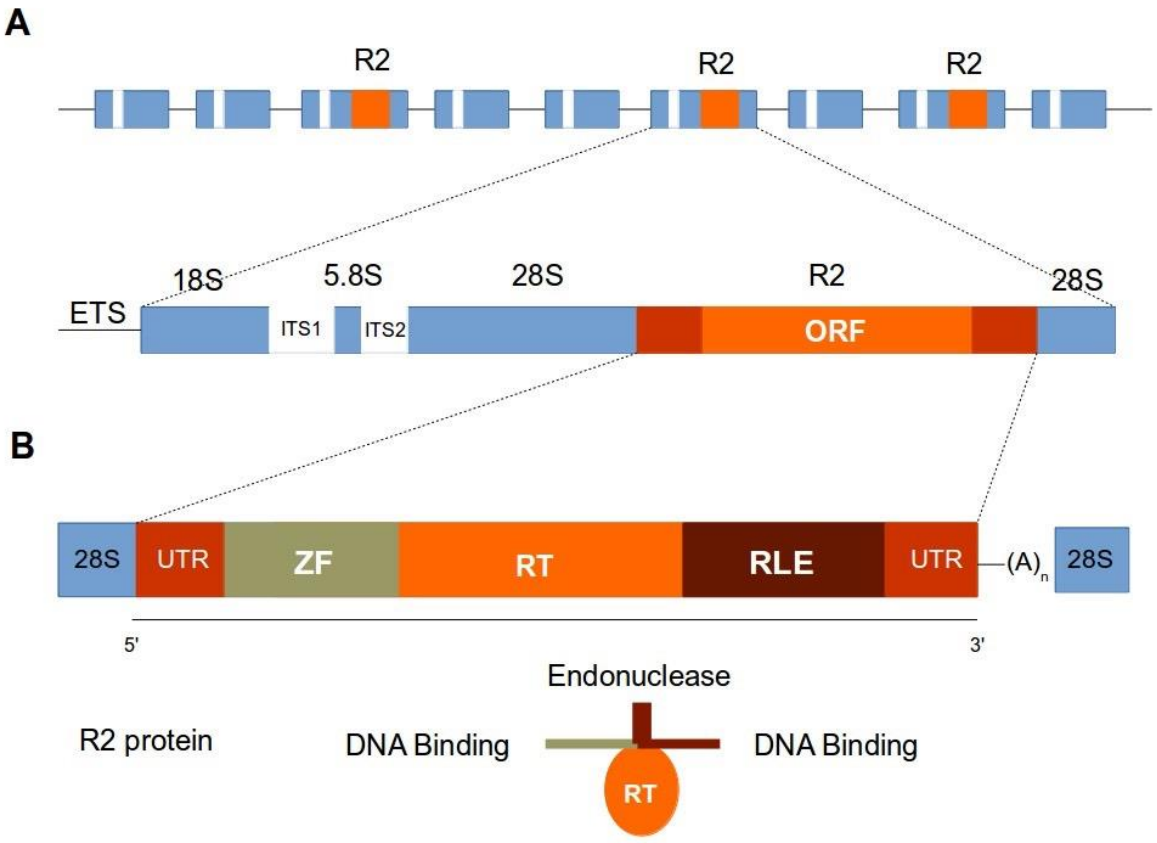


Figure 3. (**A**) The rDNA locus consists of a tandem repeat of rDNA unit, some of them are inserted by R2 element (orange boxes) which presents a single ORF flanked by two untranslated regions (UTR). The rRNA transcription unit is composed by external transcribed spacer (ETS), 18S, 5.8S and 28S genes (blue boxes), internal transcribed spacers (white boxes), and R2 insertion. (**B**) Protein domains are indicated: N-domain involved in the DNA-binding (grey box), reverse transcriptase domain (orange box) and C-terminal domain with endonuclease activity (brown box).

R2 elements have a poly(A) sequence at the 3' end of the integrated element,, because reverse transcriptase is able to add non templated nucleotides to the target site before reverse transcription of the R2 RNA template (Luan and Eickbush 1996). R2 was first identified in *D. melanogaster* (Wellauer and Dawid, 1977; Burke et al., 1987) and then in other arthropod species, vertebrates, echinoderms, flatworms and polyps. This suggested that R2 origin dates before the Radiata-Bilateria cladogenesis (Burke et al., 1993; 1999; Kojima et al., 2006).

Many phylogenetic analyses have been made over the years. These were mainly conducted on the most conserved domains, i.e. the RT and RLE (Burke et al., 1998; 1999; Kojima and Fujiwara, 2005; 2006; Luchetti and Mantovani, 2013). R2 phylogeny showed the presence of 12 sub-clades that cluster into four main clade, R2-A, R2-B, R2-C, and

R2-D following the number of N-terminal zinc-finger motifs, as showed in figure 4 (Kojima and Fujiwara 2005; Mingazzini et al., 2011; Ghesini et al.,2011; Luchetti and Mantovani, 2013).

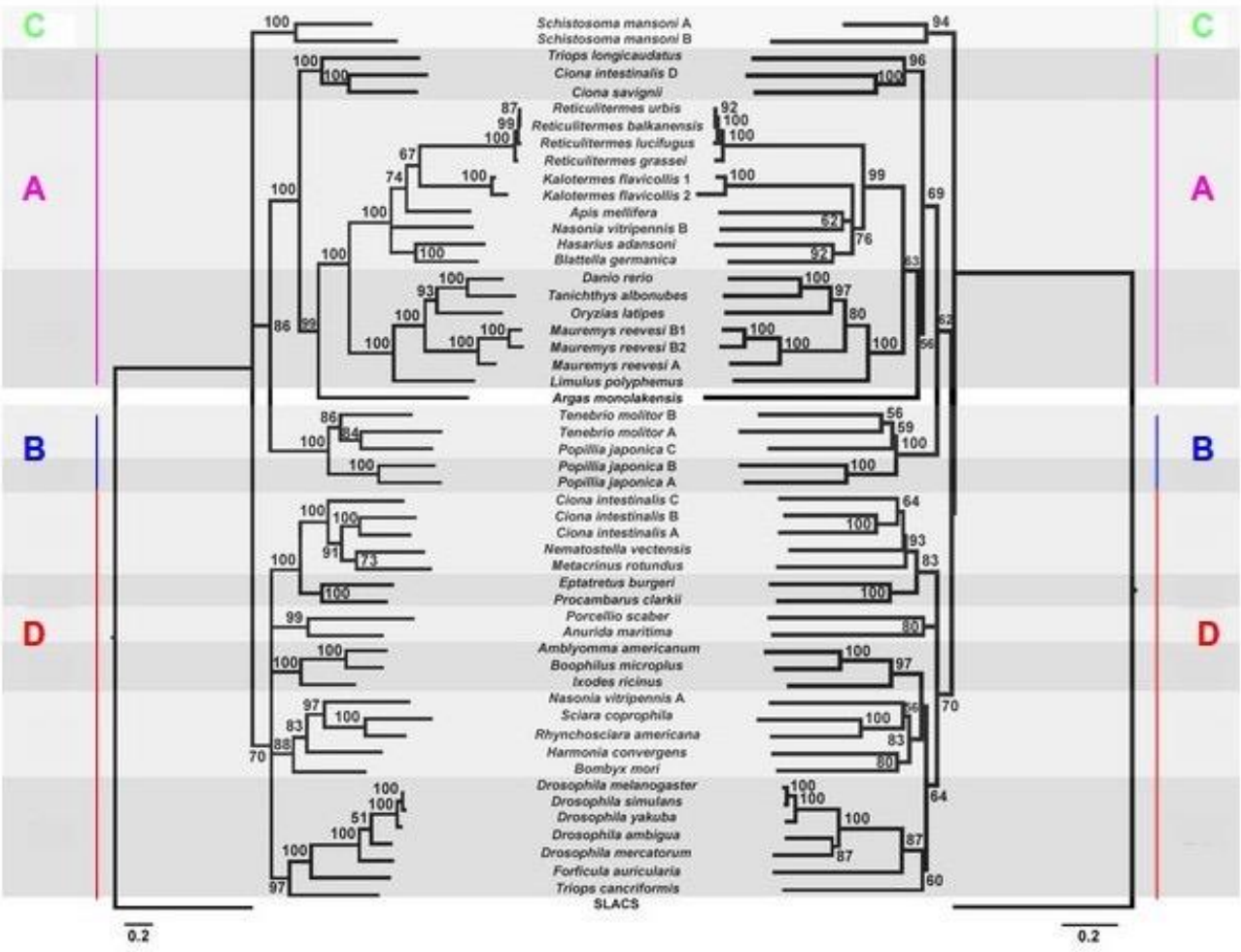


Fig. 4 Bayesian (left) and Minimum evolution (right) R2 phylogenetic trees . The number next to each node indicates posterior probability or bootstrap values, respectively. The 12 sub-clades are shown (modified from Ghesini et al., 2011).

The R2-A clade includes elements with three ZF motifs, the first ZF motif which is detected on the R2 nucleotide sequence is a CCHH motif, the second is a CCHC motif and the last one is another CCHH motif (type I, II, III, respectively). The R2-B clade was only recently shown to display the two last ZF motifs of clade A (II, III). The R2-C clade lacks the second zinc-finger of clade A (I, III). Finally, the elements in the R2-D clade have only one zinc-finger (CCHH) corresponding to the type III (Kojima and Fujiwara 2005; Luchetti and Mantovani, 2013).

1.2.2. R2 retrotranscription mechanism and ribozyme structure

As has already mentioned, retrotransposon elements move via an RNA intermediate through different mechanisms. LTR retrotransposons have promoters and regulatory regions, necessary to start/stop transcription, within their long terminal repeats. On the other hand, most non-LTR retrotransposons encode internal promoters that initiate transcription at the first base of the elements (Eickbush and Eickbush, 2010). Other non-LTR families rely on host transcription mechanism for their cotranscription. The last mechanism appears preferentially used by non-LTR retrotransposons with insertion sites within specific genes, as R2 into ribosomal genes (Eickbush e Eickbush, 2010).

The mechanism by which R2 is able to replicate and move within the ribosomal DNA, is called *target primed reverse transcription* (TPRT; Fig. 5). In the nucleus, R2 mRNA is cotranscribed with the 28S gene by host RNA polymerase I. Ribozyme activity of the R2 untranslated 5' end allows the cut and the separation of R2 mRNA from 28S. Subsequently, the transcript moves to the cytoplasm where it is translated and assembled in a polyprotein complex which, reintroduced into the nucleus, recognizes and nicks the target sequence. The polyprotein complex is composed of the full R2 transcript to which two proteins are linked, one at 3' end and another one at the 5' end. The protein at 3' end binds upstream the target site of 28S gene through a not yet identified protein domain, while the protein at 5' end binds downstream the target site through the cysteine-histidine motif and Myb. The upstream protein has an endonuclease domain and cleaves the first target DNA strand releasing the 3’ OH of the transcript. The RT domain, using the 3’ OH released, begins the reverse transcription reaction to prime cDNA synthesis. When the synthesis of the first strand is completed, the downstream subunit cleaves the sense strand; the 3' OH released is used by the RT activity as a trigger for the synthesis of the second R2 DNA strand (Christensen et al., 2006; Eickbush and Jamburuthugoda, 2008).

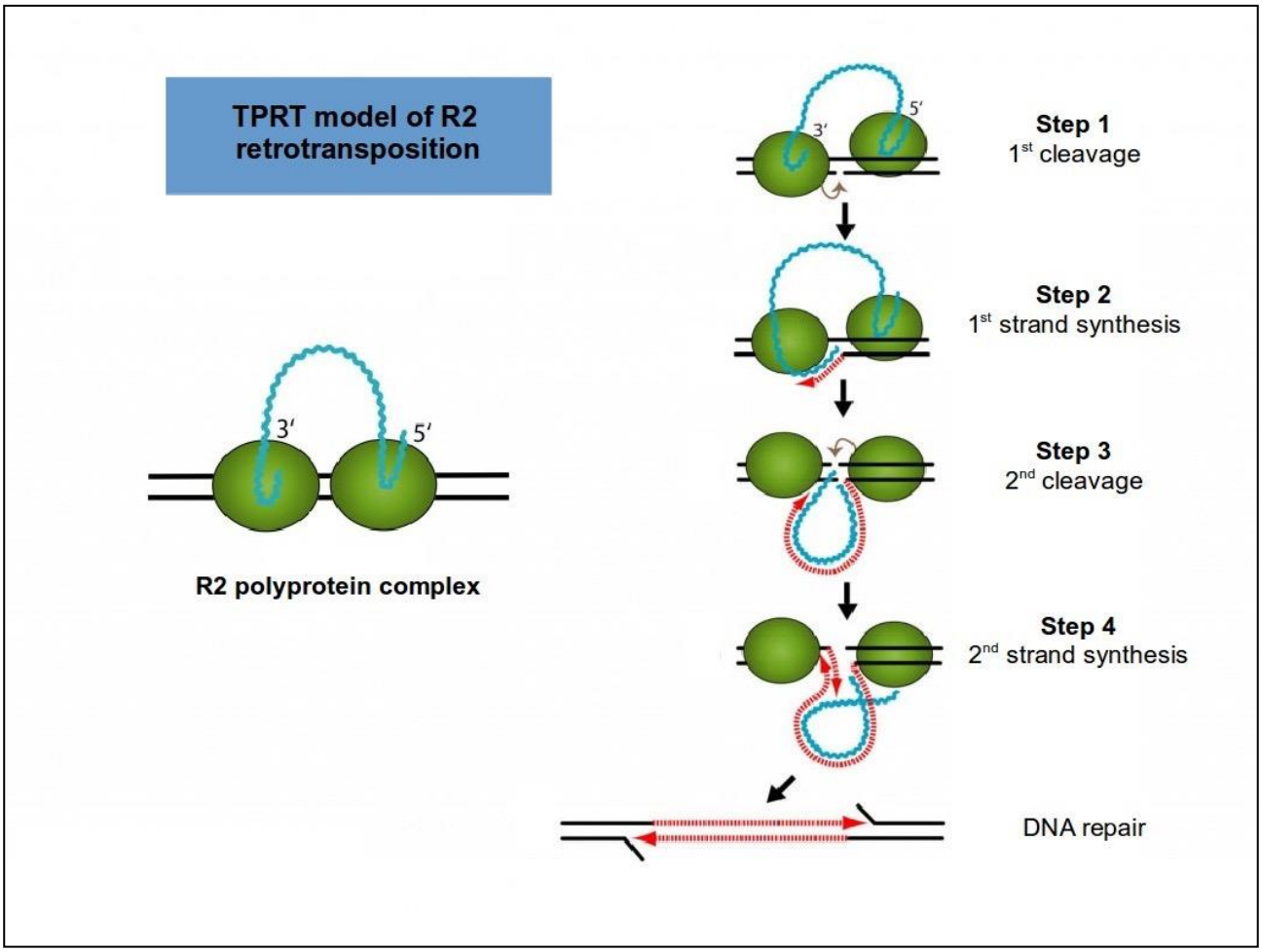


Fig. 5 Schematic representation of R2 insertion mechanism (modified from Eickbush and Jamburuthugoda, 2008).

The mechanism just presented describes the process of R2 transposition, but also it may explain why many non-LTR retrotransposons, such as R2, are characterized by conserved 3' ends and 5' ends with different lengths. This happens because the RNA template can be cleaved by cellular RNases before the end of the transcription process or, alternatively, because the reverse transcriptase can dissociate from the template before reaching the 5' end. In this ways, truncated variants are produced but their integration can still take place (Burke et al., 1993).

As described previously, R2 is cotranscribed along with the 28S gene in which it is inserted. For its transposition, R2 RNA maturation is required; this occurs through the splicing of the R2 portion from the cotranscript. Such maturation takes place through the formation of a ribozyme at the 5' end. Like many other non-LTR elements, R2 has a catalytically active ribozyme similar to the ribozyme of hepatitis delta virus (HDV; Fedor 2009; Eickbush & Eickbush 2010; Ruminski et al 2011).

The HDV virus encodes two ribozymes for its replication with similar structures:

crystallographic studies of the HDV ribozyme revealed five paired helices folded into a double pseudoknot that constrains the structure and conformation of active site. The self-cleavage reaction catalysed by the HDV ribozyme is a transesterification reaction which generates termini with 2',3'-cyclic phosphate and 5'-hydroxyl group (Ferré-D'Amaré et al.,1998).

In vitro studies demonstrated that, in several *Drosophila* species, R2 is able of self-cleavage through the formation of such a ribozyme (Ruminski et al 2011). Usually, the self-cleavage site coincides with the 5' junction of the element with the 28S gene; however, recent studies showed the possibility that the self-cleavage site can be even within the 28S gene from 9 to 36 nucleotides upstream of the R2 5' junction, so that part of the 28S gene is included in the formation of the ribozyme (Eickbush et al. 2013).

1.3. TEs survival and the host genome

1.3.1. The impact of TEs on eukaryotic genomes

TEs activity has profound effects on host genome and today many researchers agree to consider them as important actors in genome evolution.

TEs amplification/elimination can influence the genome size of many organisms. Genome size can increase also due to the amplification of different retrotransposon families in different times. Often these rapid amplifications persist for several million years. Such a long time appears surprising if we think that host mechanisms can shut down high-activity TEs families in a single generation. So far, there is no general pattern or particular property that can explain how and why some TEs can be preferentially amplified in a lineage. In fact, in some cases a particular TE family can be present in high copy number in one specie and in very low copies or missing in closed relative one. Genome size variation in the *Oryza* genus is due to polyploidization and LTR-retrotransposon proliferation (Zuccolo et al., 2007). Also in the animal kingdom there are many cases of genome size variation due to TEs. An example is provided by the giant size of the salamander genome which is rich in TEs, especially LTR elements(Sun et al., 2011).

In addition, TEs may have effects on the genome structure. They can induce chromosome breakage or rearrangements such as deletions, duplications and inversions through unequal homologous recombination or ectopic recombination (Fig. 6). An example of unequal recombination between TEs is given by LINE elements, in humans. Unequal exchange between two neighboring LINE-1 sequences in the gene encoding phosphorylase kinase beta subunit led to a DNA deletion that involved one exone of this

gene (Burwinkel and Kilimann, 1998). Ectopic recombination occurs often within a chromosome, resulting in deletion or duplication, but it also can occur between chromosomes, causing more complex abnormalities.

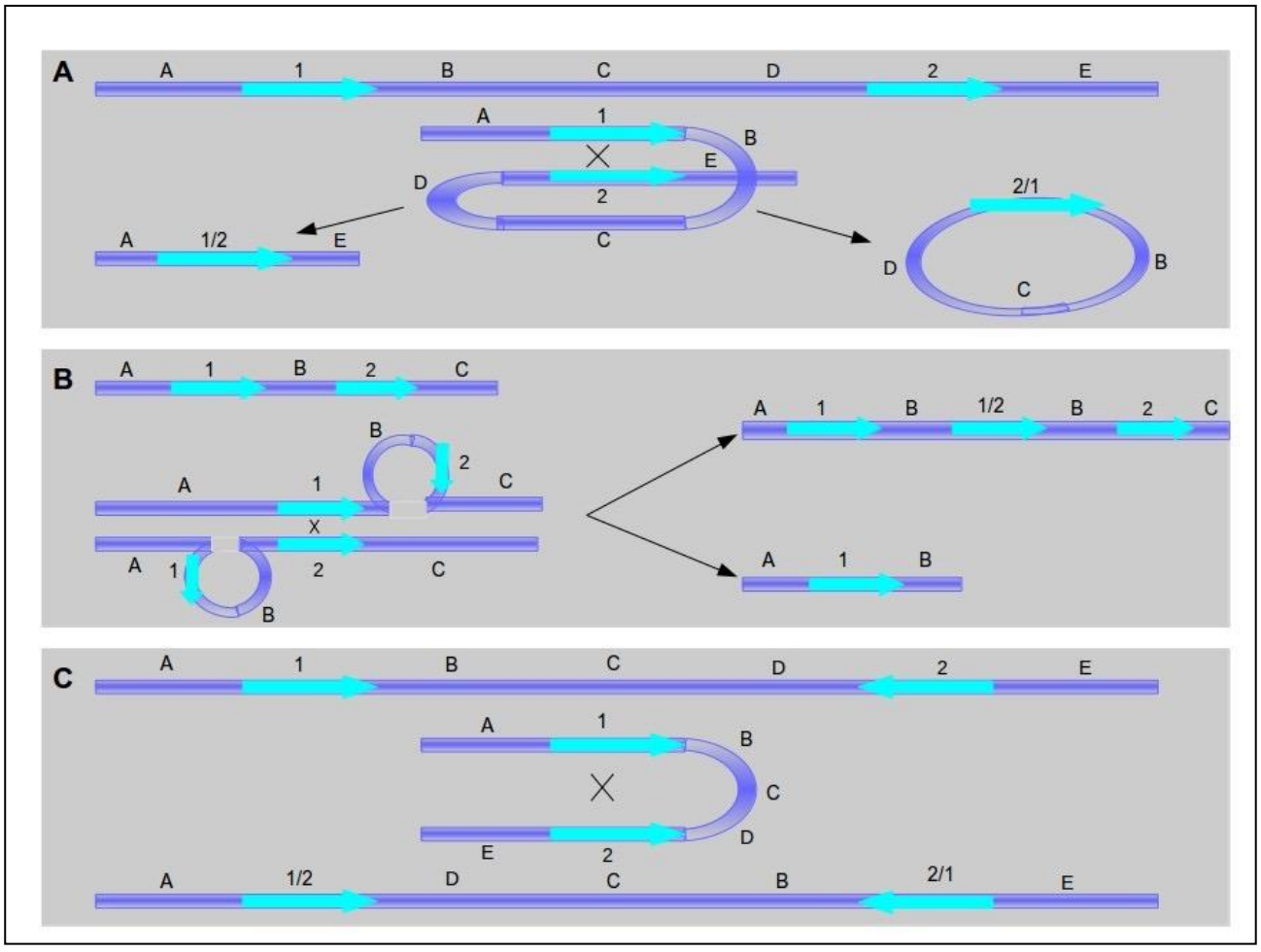


Fig. 6 Chromosomal rearrangement caused by TEs insertion. Recombination between two TEs in the same orientation on the same chromosome leads to a deletion (**A**) or a duplication (**B**), while the presence of two TEs in opposite orientations can lead to an inversion (**C**). Letters indicate coding regions, while numbers indicate TEs.

TE insertions may be considered as deleterious mutations when they occur into gene sequences, making them inactive. They may also affect gene expression when inserting at promoter and enhancer regions becoming *cis*-regulatory elements of transcription. *Cis*-regulatory elements are regions of non-coding DNA which are found near the gene they regulate, these elements functioning as binding sites for transcription factors. Several studies demonstrated that many promoters and polyadenylation signals in human and mouse genes are derived from TEs sequences (Jordan et al., 2003; Marino-Ramirez et al., 2005). Some retrotransposons can introduce intragenic polyadenylation signals, such as human L1 and *Alu,* leading to the production of more than one transcript from a single gene in a process similar to alternative splicing (Cordaux and Batzer, 2009).

TEs can also be subject to “exonization”. Exonization takes place when a TE inserted in the untranslated regions of a gene is rescued as a new exon. The process is often associated with alternative splicing events. In addition, TEs insertions can also deliver novel introns (Sela et al., 2010; Chénais et al.,2012).

Another TEs contribution on the evolution of protein coding region derives from “molecular domestication” which occurs when TEs encoded proteins or domains become co-opted into functional host proteins. Several cases of host proteins derived from the complete coding sequence of the transposase have been reported. One famous example of TE domestication is the case of *transib* DNA transposon that causes the unique mechanism of genetic recombination that occurs only in developing lymphocytes (V(D)J system; Kapitonov and Jurka, 2005).

Cis-regulatory elements, exonization/intronization and molecular domestication are phenomena also known with the term exaptation. The term exaptation was coined by Gould and Vrba (1982), to refer to those characters having a particular function and co-opted for a new use.

1.3.2. TEs transposition rate dynamics and host reproductive strategies

Initially, TEs were considered genomic parasites due to their ability to reduce the fitness of host organism like most deleterious mutations, but, unlike deleterious mutations, they are able of independent activity (Dawkins, 1976; Doolittle and Sapienza, 1980; Orgel and Crick, 1980). As described above, TEs activity outcome in the host genome are not always deleterious, the majority of insertions being nearly-neutral. However, an overload of TE insertions is certainly deleterious for the host. It is not therefore surprising that host organisms have evolved various mechanisms to prevent TEs proliferation, such as DNA methylation, that inhibits TE transcription, and the presence of small RNAs that can inhibit TE mobility through the post-transcriptional disruption of transposon mRNA (Levin and Moran, 2011).

Another mechanism which influence TEs activity is the host reproductive strategies. The strategic role of sexuality in TEs evolutionary dynamics was suggested by Hickey (1982). He formulated the hypothesis that bisexual reproduction, through homologous chromosomes recombination and reassortment during meiosis, can control the spread and proliferation of mobile elements, while unisexual reproduction would lead to an increase of these elements density due to the inability to eliminate them (Arkhipova & Meselson 2005; Dolgin & Charlesworth 2006). These relationships are well-described by two evolutionary

hypothesis: the Red Queen and the Muller's ratchet hypotheses (Salathé et al., 2008). The first models the arm race established between TEs and the host genome, whose coexistence implies an equilibrium between TE replication rate and elimination through recombination and selection. The second hypothesis postulates that organisms with no recombination (or with non-effective recombination) will accumulate deleterious mutations (TEs, in this instance) until their extinction.

The Bdelloidea rotifers are the only organisms which have undergone successful radiative evolution in an ancient unisexual condition. First studies showed that these organisms had only DNA transposons, while retrotransposons appeared totally absent. These evidences suggested the idea that Bdelloidea could tolerate elements which move from one site to another through the excision of preexisting copy, than elements producing copies of themselves, thus increasing their content in the genome (Arkhipova & Meselson, 2000). More recently, though, the presence in these organisms of *Athena* non-LTR retrotransposons in telomeric and subtelomeric regions (and their absence in gene-rich region) was reported (Gladyshev and Arkhipova, 2010). Detailed analyses on TE content and dynamics in taxa with different reproductive strategies suggested that reproductive strategies may affect the inheritance of some specific elements, at least in *Aradopsis thaliana* (Wright et al., 2001; Lockton and Gaut, 2010) and *Caenorhabditis* spp. (Dolgi et al., 2008).

On the other hand, the study conducted by Kraaijeveld et al. (2012) on sexual and asexual populations of the parasitoid *Leptopilina clavipes* genome suggested that the density of TEs depends on the families considered rather than from reproductive strategies. The genomic sequencing of the bisexual and parthenogenetic populations showed the presence of transposons belonging to the main classes, in both populations. Rather, the bisexual populations have more non-LTR elements, while the parthenogenetic strains have more DNA transposons. In this work, a general copy number increase for all classes of TEs, as predicted by classical theory, has not occurred in asexual populations. Very recently, an additional analysis on sexual/asexual pairs of arthropod taxa, reinforced the idea that reproductive strategies impact very little on genomic TE content (Bast et al., 2015). Thus, the role of reproductive strategies on TE proliferation and evolution remains debated.

Whatever the forces acting on TE load, different mechanisms evolved that allowed transposable element to survive and increase in copy number in the host genome. Here I will detail only two mechanisms.

1.3.3. Site-Specific insertions

While many TEs can insert into different sites, some TEs exhibit a target site selection. Target site selection appears as a strategy to optimize the element-host relationship. Many TEs have evolved to target specific regions in the host genome, where their activities would cause minimal damage to the host (Levin and Moran, 2011). For example, some elements, including the yeast Ty1 and Ty3 elements and the *Drosophila* P element, usually insert upstream of promoters, thus decreasing the probability of element insertion into coding sequences. In bacteria, transposon Tn7 has the capacity to utilize different kinds of target sites, either a single neutral chromosomes site or mobile plasmids (Peters J.E. 2014). Another niche that allows TE insertions with limited damage to host genome is represented by multiple copy genes. This is the strategy used by R elements. The ribosomal DNA of eukaryotes (DNA), in fact, is composed of tandem units repeated from hundreds to thousands of times. Each unit is represented by three genes (18S, 5.8S and 28S) flanked by spacer sequences; upstream the 18S gene one external transcribed spacer (ETS) is present, while the 5.8S gene is separated from the 18S and 28S genes by two internal transcribed spacers (ITS1 and ITS2, respectively). A not transcribed intergenic spacer separates each unit (IGS; Nei e Rooney, 2005). The sequences of the ribosomal genes and spacers evolve in a concerted manner (Nei e Rooney, 2005; Eickbush e Eickbush, 2007). Concerted evolution (Fig. 7) is achieved through molecular drive, a dual process independent from natural selection and genetic drift. This process is first due to a number of turnover genomic mechanisms (gene conversion, unequal crossing-over, rolling circle replication, transposition etc.), that determine the spread of new gene variants within the same genome (homogenization), while the spreading in the population occurs through bisexual reproduction (fixation). Molecular drive, therefore, drives the concerted evolution of a locus, generating a greater similarity of repeated sequences within a taxonomic units (population, species etc.) than among different ones (Ganley et al.,2007).

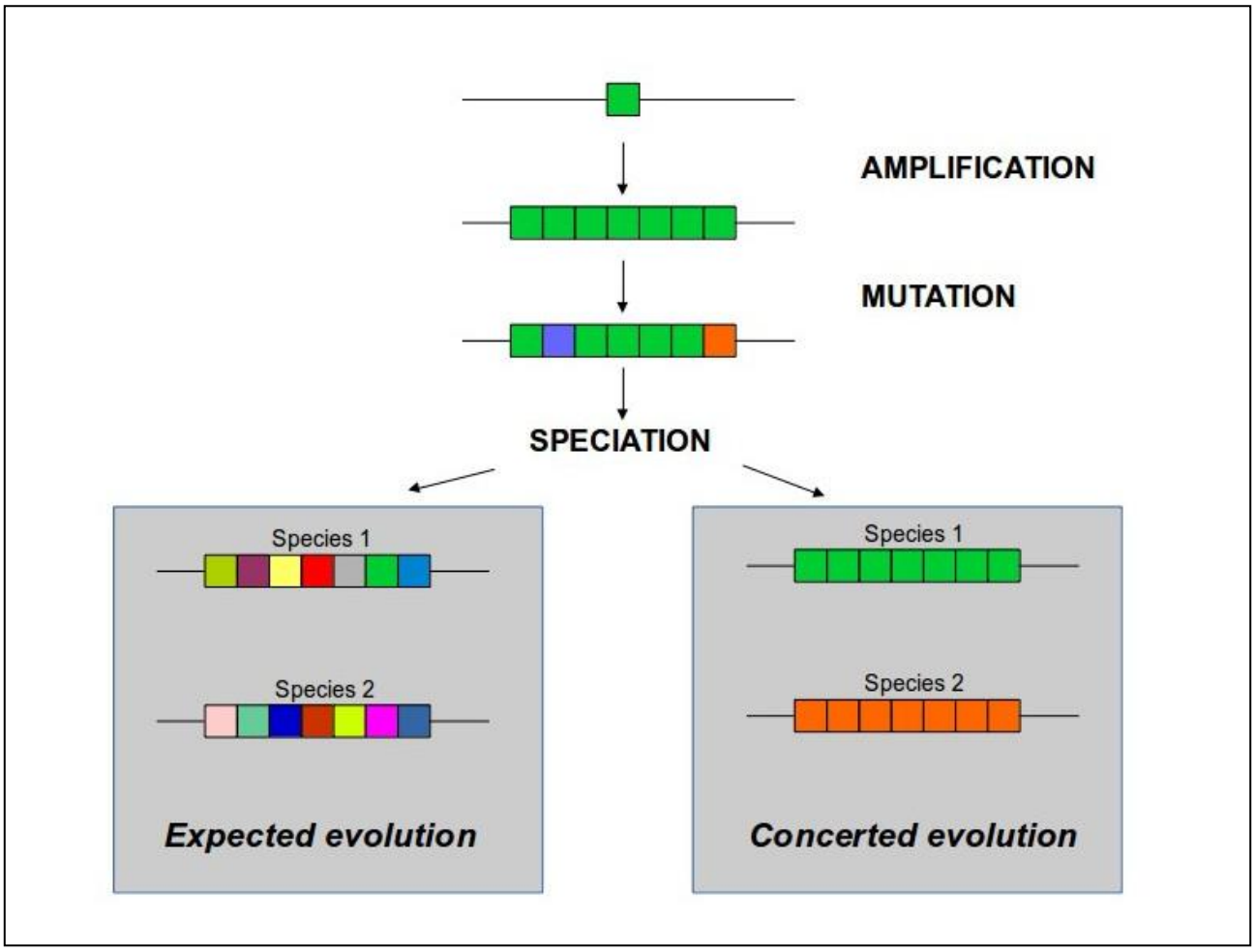


Fig. 7 Expected evolution *Vs* Concerted evolution.

Consequently, also R1 and R2 insertions dynamics appears influenced by the peculiar evolution of its niche. In fact, studies of the R1 and R2 elements in *Drosophila* (Perez-Gonzalez and Eickbush, 2001) have shown that new insertions are rapidly eliminated, because there is a strong selective pressure against inactive rDNA units which are eliminated often by unequal crossover between sister chromatids, and then replaced by new rDNA units through compensatory genomic turnover mechanisms. The elimination of 28S inserted units and their substitution with new functional 28S units can be advantageous but this also leads to new sites for further R2 insertions (Eickbush e Eickbush, 2007).

1.3.4. Horizontal transfer

In the early 1990, a study on the P element of *Drosophila* (Daniels et al., 1990) revealed that this element was able to colonize new genomes by means of horizontal transfer (HT). This discovery suggested an additional way which enable TEs to escape host mechanisms that contrast their proliferation. HT is the transfer of genetic material between

reproductively isolated, even distantly related taxa leading to very different evolutionary dynamics. In standard vertical inheritance, i.e. genome transmission from parents to offspring, TEs are subject to host suppression mechanisms, aimed at limiting their mobility and their expansion in copy number, and also the stochastic loss by chromosome segregation. HT allows TE invasion of a new genome, giving fire to a new arm race in the light of the Red Queen hypothesis (Salathé et al., 2008). It is to be noted that with the increasing number of available genomes (also from non model organisms), more and more cases of horizontal transposon transfer (HTT) have been identified, leading to the hypothesis that HT could be an essential part of the life cycle of some types of TEs.

HT is apparently more frequent among Class II than Class I TEs, this being possibly linked to the different transposition mechanisms used by the two classes. DNA transposons have a more stable double-stranded DNA intermediate, while retrotransposons have in general terms a relatively unstable RNA intermediate that is reverse-transcribed directly into the chromosomal target site, so that the possible transfer outside the cell nucleus is limited. A notable exception is given among Class I elements, by a small group of LTR retrotransposons (as *Gipsy* elements) similar to retroviruses in terms of replication mechanism and structural organisation (Malik et al., 2000). These elements have two open reading frame, corresponding to retroviral *gag* and *pol* genes, and an extra ORF in the same position as the *env* gene found in retrovirus genomes. This ORF consist of a putative transmembrane domain at the C-terminus, multiple putative N-glycosylation sites, and putative protease cleavage sites, similar to the cleavage sites in a variety of retroviral *env* proteins, at conserved positions (Song et al.,1994). The protein encoded by this ORF is comparable for structure and function to the envelope protein present in retroviruses. For these reasons, these LTR retrotransposons can be able to move easily between species by infecting new cells (Havecker et al., 2004). Without virus-like envelope proteins, other TEs require a vector to facilitate HT.

On the whole therefore, DNA transposons and LTR retrotransposons appear prone to HT (Schaack et al., 2010). However, some new cases of HT concerning non-LTR retrotransposons have also been found recently: CR1 clade in *Maculinea* butterflies and *Bombyx* (Novikova et al., 2007); L1 and BovB elements among vertebrates (Ivancevic et al., 2013); Sauria SINE in snakes and rodents (Piskurek and Okada, 2007) and R4 retrotransposon between *Aedes* and *Anopheles* (Biedler et al., 2015).

The exact mechanism allowing HTT is still unknown. The hypothesized routes (as illustrated in figure 8) range from very simple ones, such as a direct transfer without or with one intervening vector (Fig. 8a-d), to more complex systems involving more than one

vector (Fig. 8e-f; Schaack et al., 2010). Some LTR retrotransposons can produce virus-like particles (VLPs) able to infect the germ line and to be potentially transmitted as intracellular virus-like particles without the need of any independent vector (Fig. 8a), but vectors such as bacteria/viruses/parasites/parasitoids (Fig. 8b-d) can be called into action in HTT. Mites do not need to integrate necessarily the sequence transferred into their own genome but they can transfer TE directly from the donor to the host. Also intracellular symbiotic bacteria can be good vectors, in particular *Wolbachia* and spiroplasms since they live inside germ cells. These endosymbionts can only be a shuttle for TE or become proper vectors (Fig. 8d). It has been also demonstrated that bacteriophage infection is a common feature of *Wolbachia,* so bacteriophage could be an intermediate vector to transfer TE to intracellular symbiotic bacterial genomes which would become the final vector (Fig. 8e) or maybe intracellular symbiotic bacterial can be intermediate vectors transferred between different host species by parasites or parasitoids (Fig. 8f; Loreto et al., 2008).

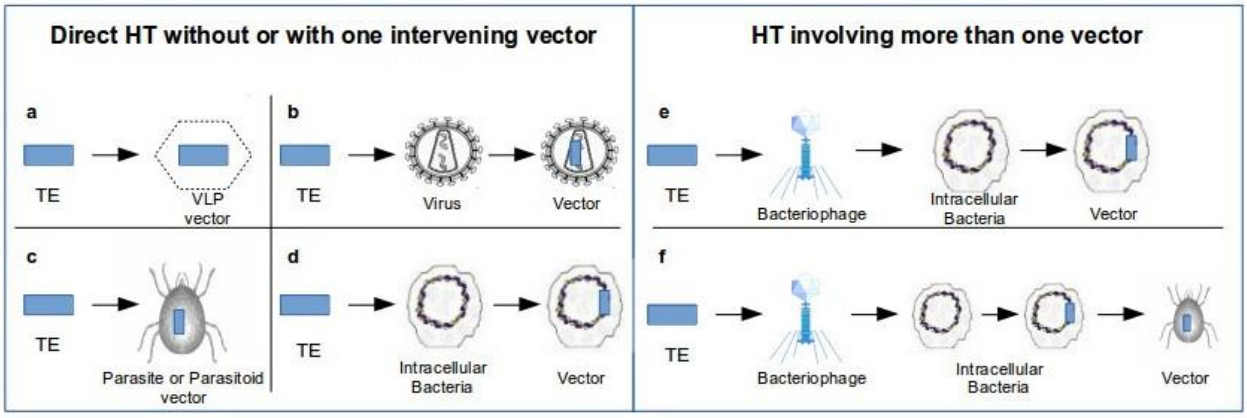


Fig. 8 Different horizontal transfer mechanisms and vectors that have been suggested: (a) TEs produce virus-like particles (VLPs) that may work as a vector; (b) TEs could be transported by DNA virus; (c) TEs transfer through parasites and parasitoids; (d) intracellular symbiotic bacteria as vectors; (e) bacteriophage as intermediate vector to transfer TE to final vector intracellular symbiotic bacterial; (f) the intracellular bacteria as a further intermediate vector and parasites or parasitoids as a final vectors (modified from Loreto et al., 2008).

Generally, possible HT events are inferred through three criteria (Silva et al.,2004; Loreto et al., 2008), i.e. when the following is observed:

I) TE from different host species with sequence similarity higher than that expected on the basis of the hosts divergence;

II) incongruences between TE and host phylogeny;

III) patchy distribution of the TE within a group of taxa.

The first method to infer HT is based on sequence similarity comparisons, allowing detection of elements with a high degree of sequence similarity in divergent taxa. The second method tries to identify if the TE distribution in its host species is not consistent with host phylogeny. When TEs are transmitted vertically, in fact, their phylogenetic history is expected to retrace, at least in broad terms, that of their hosts. Another line of evidence for horizontal transfer is the so-called “patchy” distribution of a TE family among closely related taxa; in particular it refers to the presence of a TE in a given lineage and its absence in the sister lineage (Silva et al., 2004). A HT case can be confirmed when all three approaches give consistent results.
HTT is an important aspect that may play a major role in the observed diversity and abundance of eukaryotic TEs.

1.4. The order Phasmida

The order Phasmida comprises heterometabolous, phytophagous and nocturnal insects that reside in almost all temperate and tropical ecosystems. This order includes a little more than 3000 species (Brock, 2011). The term Phasmida originated from the Latin word "phasma" which means ghost; in fact, the most evident characteristic of this order is the cryptic mimicry. The morphology of these insects and their movements resemble the form and movement of branches and leaves moved by wind making them difficult to spot. This is why they are commonly known as stick or leaf insects (Fig. 9). They have a small and mobile head with compound eyes and masticatory mouthparts. Moniliform antennae can be very short or longer than the body. Mesothorax and metathorax are well developed; the first abdominal segment is fused with the metathorax forming the "median segment", a taxonomic character of the order. Generally, the insects belonging to this order are wingless and have long legs that allow them to climb and walk on the feeding plant.
Their reproduction can be both/either bisexual and/or unisexual. When both sexes are present, the animals show sexual dimorphism: apart from genitalia, the male is usually smaller and slender than the female. The male genitalia are covered by the ventral region of the ninth segment called *poculum,* while female genitalia are covered by the ventral region of the eighth segment, called *operculum.* The eggs (Fig. 10), resembling seeds, have the anterior side closed by an operculum, while the dorsal side is differentiated in a micropylar plate (Sellick, 1998).

Fig. 9 *B. rossius* on its feeding plant, *Rubus spp*. (photo - Francesco Martoni).

Fig. 10 *Bacillus rossius* eggs.

1.4.1. The *Bacillus* genus

The genus *Bacillus* Berthold, 1827 is one of the four genera present in the Mediterranean area (Fig. 11). The insects belonging to this genus are wingless with a body 6-10 cm long. *Bacillus* taxa have been characterized at the morphological, allozymatic and molecular levels (Mantovani et al.,2001). Morphological analyses took into consideration adult and egg characters. The basal species of the *Bacillus* genus are: *B. rossius, B. atticus* and *B. grandii*.

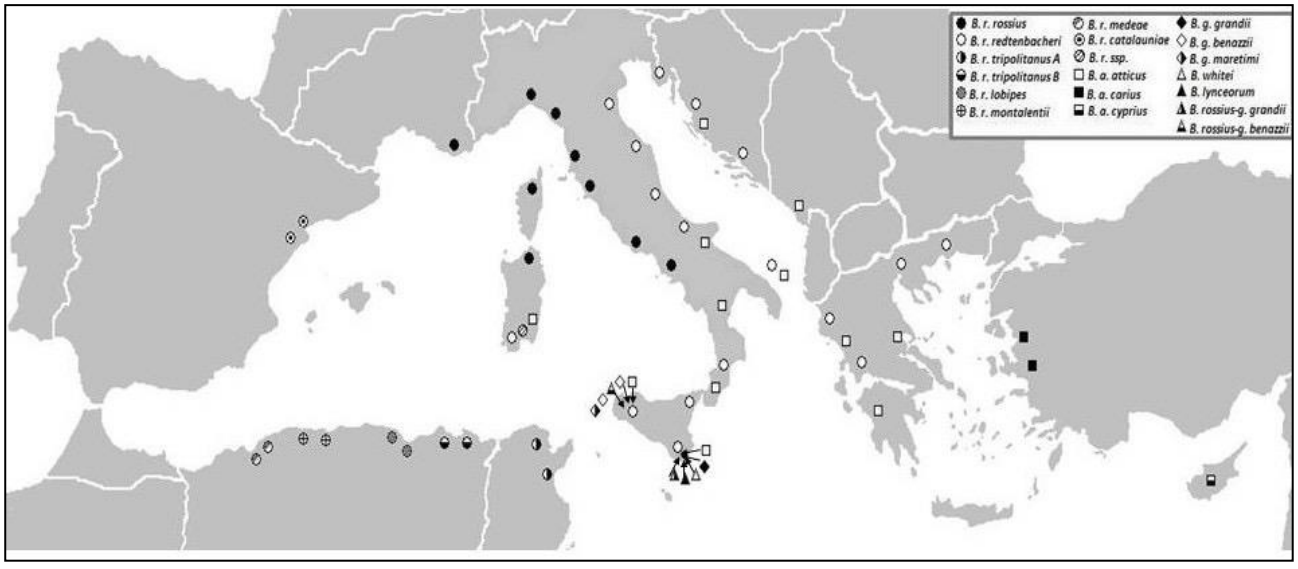


Fig. 11 *Bacillus* taxa distribution in the Mediterranean area.

B. rossius is widespread along the central-western coasts of the Mediterranean Sea, with eight subspecies mainly recognized on genetic distances derived from the analyses of the enzyme-gene systems: *B. rossius rossius, B. rossius redtenbacheri, B. rossius tripolitanus A, B. rossius tripolitanus B, B. rossius lobipes, B. rossius montalentii, B. rossius cataluniae* and *B. rossius medeae*.

In Italy only the first two subspecies are present both with gonochoric and parthenogenetic populations: *B. rossius rossius* Rossi, 1790 all along the Thyrrenian coast (Liguria, Tuscany, Latium, Campania) and in Sardinia (Fig. 11), and *B. rossius redtenbacheri* Padewieth, 1899 along the Adriatic and Ionian coasts, in Sicily, the Aeolian Islands and a small area in south-eastern Sardinia (Scali et al. 1992; Fig. 11).

B. grandii is also a bisexual species but, at variance of *B. rossius*, it is strictly amphigonic. The species is endemic in the Sicilian area and it is differentiated in three subspecies: *B. grandii grandii* Nascetti and Bullin, 1981 with some relict populations in the Iblean region, *B. grandii benazzii* Scali,1989 present in a limited area of the north-western Sicilian coast and in the Levanzo Island (Aegadian Archipelago) and *B. grandii maretimi* Scali and Mantovani, 1990 endemic of Marettimo Island (Aegadian archipelago; Fig. 11).

B. atticus is extensively distributed in the central-eastern Mediterranean area and it comprises a complex of diploid and triploid thelytokous obligatory parthenogenetic populations. It is divided into three subspecies depending on allozymes and karyological characterisation: the diploid *B. atticus atticus* Brunner, 1882, the most common taxon in the area; *B. atticus carius* Mantovani and Scali, 1985, with triploid populations in Greece and in Turkey and a diploid one in Turkey; *B. atticus cyprius* Uvarov, 1936, endemic of Cyprus Scali et al. 1995 (Fig. 11). Although allozymatic analyses and Nei's genetic distances suggest that *B. atticus cyprius* has a higher affinity with *B. atticus atticus,* body and egg morphology and karyological characterisation seem to support a subspecific differentiation (Mantovani et al., 1995).

In the Sicilian area, four hybrid lineages derived from crosses of *B. rossius redtenbacheri*, *B*. *atticus atticus* and *B. grandii grandii* or *B. grandii benazzii*: the diploid *B. whitei* Nascetti and Bullini, 1981 (B. *rossius redtenbacheri*/B. *grandii grandii*) and the triploid *B. lynceorum* Nascetti and Bullini, 1982(B. *rossius redtenbacheri*/B. *grandii grandii* /B. *atticus atticus*) with an obligatory parthenogenetic reproduction, and the diploid *B. rossius-grandii grandii* and *B. rossius-grandii benazzii* which are the first cases of natural hybridogenesis and androgenesis among invertebrates (Mantovani et al., 1990; Tinti et al., 1995).

Bacillus taxa show morphological diagnostic characters such as the eye's pigmentation (Fig.12), the female subgenital plate and the pattern of the egg corion (Mantovani et al., 1992).

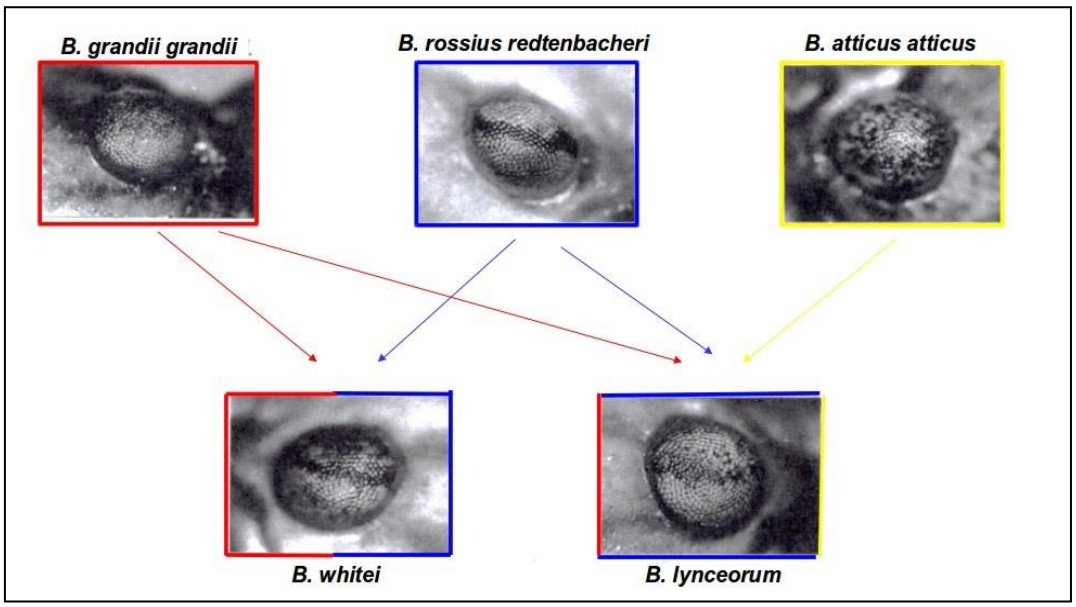


Fig.12 Eye pigmentation of the parental and hybrid taxa of the *Bacillus* genus (modified from Mantovani et al. 1992)

B. rossius exhibits a well-defined pigmentation (Fig.12), which forms a bar in the middle of the eye, and the egg corion seems like as an irregular lace-network of ribbons. *B. grandii grandii* shows a totally pigmented eye without bar (Fig.12), the smooth egg capsule is covered by a rather regular lace-network of ribbons. *B. atticus* shows a pigmentation uniformly distributed around the eye and a corionic pattern closely resembling that of *B. grandii grandii*. The hybrid eye pigmentation appears as a mix of parental eye pigmentation (Fig. 12). Capsule sculpturing of *B. whitei* appears to be given by an irregular network of indented sharp ribbons while capsule pattern of *B. lynceorum* is rather variable according to the specimen's origin, the lace-network of sharply raised ribbons being much more regular than that of *B. whitei* (Mantovani et al, 1992).

1.4.2. Reproductive strategies

B. rossius is a facultative parthenogenetic, i.e. it shows in different part of its range, either parthenogenetic or bisexual populations.

Parthenogenesis is a non canonical reproduction in which the eggs will have a virginal development, giving organisms characterized only by the maternal genetic information.

B. rossius has an automictic parthenogenesis (Fig.13) in which the maturation of the egg cells occurs through a normal meiosis. The segmentation of the haploid germ cell continues until the formation of thousands of nuclei. At this point just some cells, through anaphase restitution, restore the diploid condition and carry on the embryonic

development. This mechanism produces polyclonal offspring, which are homozygous at all loci just after one generation (Scali et al., 2003). *B. rossius* parthenogenesis is thelytokous because it produces only female offspring.

Also *B. atticus* reproduces by an automictic parthenogenesis but with a mechanism different from *B. rossius* (Fig. 13). In diploid forms during a normal first meiotic division there is the formation of bivalent chromosomes and the segregation of homologous chromosomes. The nuclei are fused at prophase II, restoring the diploid condition. During the second meiotic division, one of the two nuclei degenerates while the other will generate the embryo (Scali et al., 2003). In the triploid forms the meiotic mechanism is essentially the same. During the first meiotic division, in addition to the formation of bivalent chromosomes, there is the formation of univalent and multivalent chromosomes, consequently the nuclei produced are not balanced. The normal triploid condition is restored with the fusion of the products of the first meiotic division. These products undergo a second division that produces a degenerative polar globule and a triploid nucleus perfectly balanced (Scali et al., 2003). During the first meiotic division the homologous chromosomes can recombine. This peculiarity of the automictic mechanism in *B. atticus* elucidates the heterozigosity reduction in the offspring of heterozygous females, and is also at the base of the complex clonal structure of the population of such species (Scali et al. 2003).

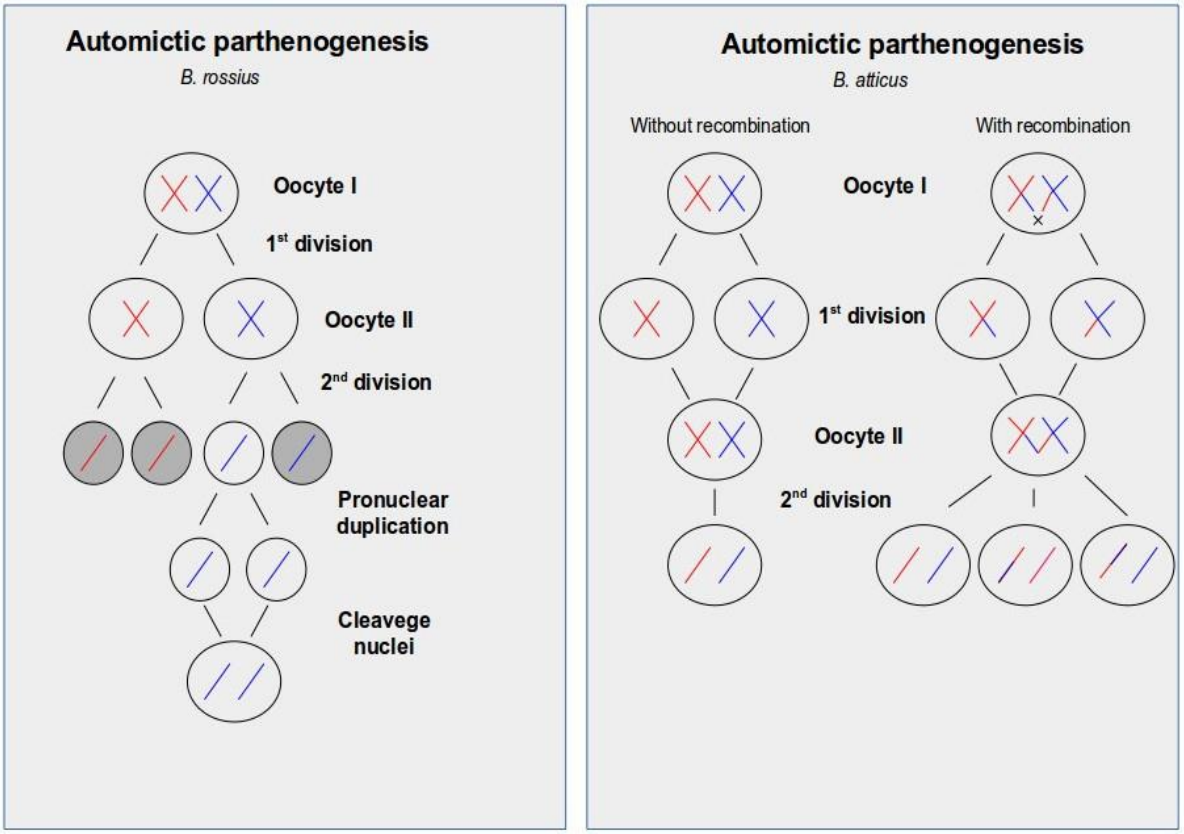


Fig. 13 Schematic representation of automictic parthenogenesis in *B. rossius* and *B. atticus*.

The parthenogenesis accomplished by the hybrids *B. whitei* and *B. lynceorum* is an apomictic one and meiosis appears suppressed, resulting in the invariant transmission of many fixed heterozygous loci of the mother to progeny. During prophase I, in fact, a DNA extra-synthesis occurs, with the formation of tetrachromatidic chromosomes. The two meiotic divisions produce therefore an unreduced pronucleus plus three degenerating polar bodies. (Mantovani et al.,1992; Scali et al., 1995).

Unlike *B. atticus,* in which the automictic mechanism, through recombination, generates some variability in the offspring, the apomictic mechanism maintains the genetic structure of the parental females.

The hybrid lineages *B. rossius-grandii grandii* and *B. rossius-grandii benazzii* reproduce by hybridogenesis and androgenesis. In the hybridogenetic female germ line the elimination of the paternal genome takes place and only a maternal haploid chromosome set (of *B. rossius redtenbacheri,* see paragraph 1.4.3.) is passed to the offspring in a hemiclonal way. This allows to overcome the meiotic constraints of heterospecific chromosome pairing. In the same hybridogenetic lineages of *Bacillus* it was observed, for the first time in the animal kingdom, the androgenesis phenomenon in which the embryo derives from the paternal nuclear genome only, having eliminated the maternal pronucleus. The formation of a diploid nucleus takes place by fusion of two male pronuclei.

This reproductive strategy, discovered for the first time in the Animal Kingdom in *Bacillus*, was thereafter described also for *Clonopis* and *Leptynia* genera, members of the Phasmida order, and in the *Corbicula* genus (Ghiselli et al. 2007; Milani et al, 2010; review Grebelnyi, 2009).

1.4.3. Phylogenetic relationships

The phylogenetic relationships among *Bacillus* taxa have been clarified through morphological, allozymatic, mitochondrial cytochrome oxidase II and satellite DNA analyses. All data obtained from these analyses are in agreement between them. Allozimatic analyses reveal a minor divergence level in the clade *B. grandii* (distance for Nei: D= 0.21-0.26) , a more important differentiation between *B. grandii* and *B. rossius* (D= 1.23-1.50) and a higher similarity between *B. atticus* and *B. grandii* (D= 0.31-0.36) than *B. rossius* (D= 1.48-1.63) (Scali et al. 1995).

The phylogenetic analysis performed on the mitochondrial cytochrome oxidase II (COII) gene confirmed the previously inferred relationships and further allowed to prove that all the unisexual hybrids of the genus are the result of a series of independent events of

hybridization in which the subspecies *B. rossius redtenbacheri* is always the maternal ancestor (Fig.14; Mantovani et al., 1999; Mantovani et al, 2001).

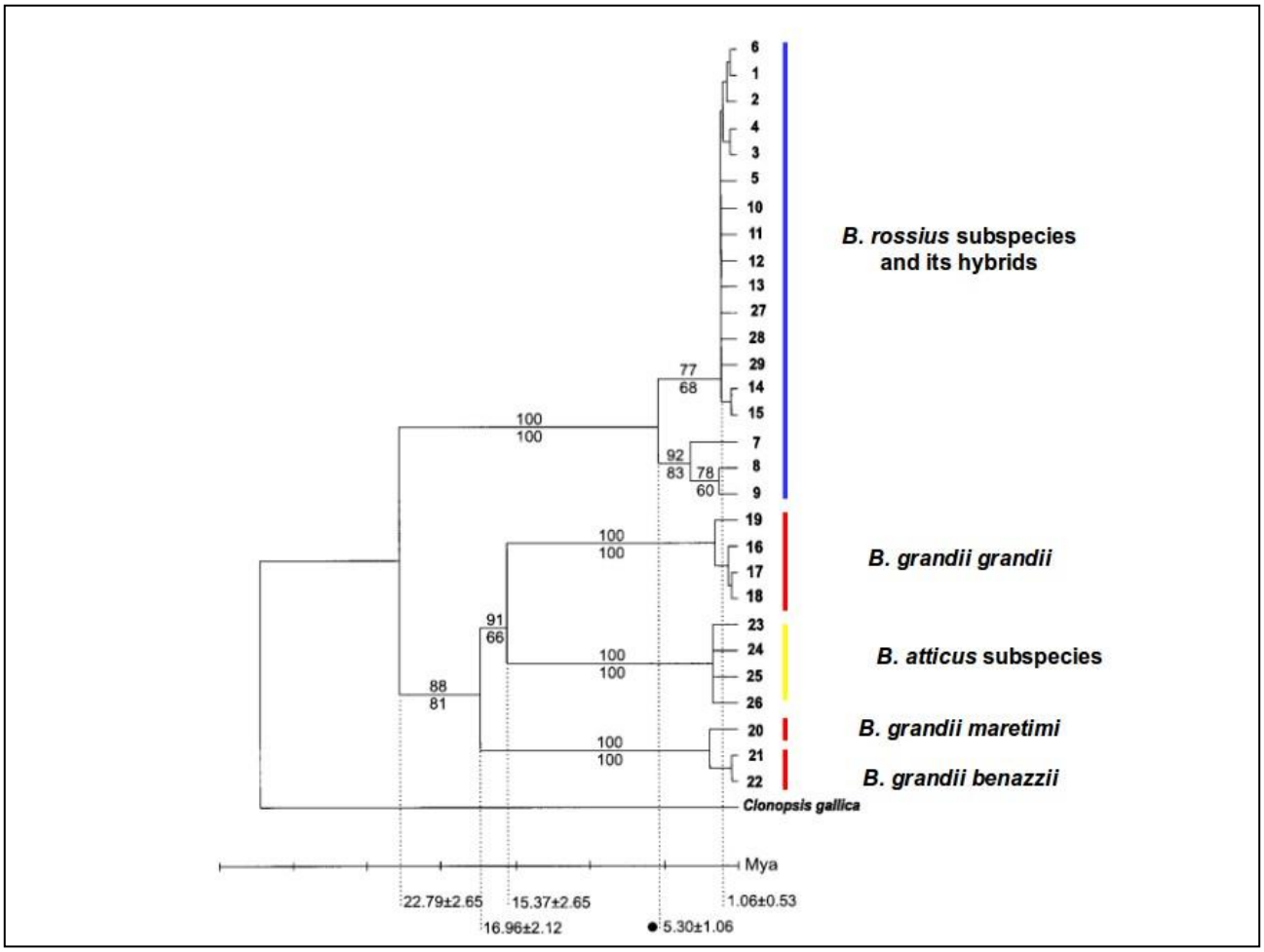


Fig.14 Phylogenetic relationships of the taxa of the *Bacillus* genus based on Maximum Parsimony and Neighbor-Joining methods (modified from Mantovani et al, 2001).

The tree (Fig.14), produced on COII sequences, in fact, showed two distinct clusters: the first includes haplotypes attributable to *B. rossius*, *B. whitei*, *B. lynceorum*, *B. rossius-grandii benazzi* and *B. rossius-B. grandii grandii* androgenetic haplotypes; the second one includes the haplotypes of *B. grandi*i and *B. atticus*.

The phylogenetic relationships of *Bacillus* taxa based on allozymatic and mitochondrial analyses are confirmed in the analysis of a satellite DNA family, called Bag320 (Luchetti et al. 2003). Luchetti et al. also demonstrated that the variability of the satellite DNA follows a concerted evolution model and depends on the host reproductive strategies. The comparison between *B. grandii* subspecies showed a significant differentiation of repeated sequences ($p < 0.001$), while no significant differences were found between populations of the same subspecies. The *B. atticus* unisexual populations instead showed a variability of the satellite DNA equally distributed at all considered levels.

1.5. The order Notostraca (Branchiopoda, Crustacea, Arthropoda)

The order Notostraca comprises branchiopod crustaceans in which the head (cephalon), the thorax (pereion) and the anterior part of the abdomen (pleon) are enveloped by a shield-like carapace. Resembling somehow a tadpole, they are commonly known as tadpole shrimps. This order comprises only one family, Triopsidae, and two genera, *Triops* Schrank, 1803 and *Lepidurus* Leach, 1816.

Body ranges in length from one to ten centimetres. The carapace shows at the hind edge a semicircular posterior emargination, the sulcus. Some species have spiny carapace carrying the odd naupliar eye, sessile compound eyes and a dorsal sensorial organ. The cephalon presents two pairs of antennae, a pair of mandibles and two pairs of maxillae. The pereion is composed by eleven thoracic segments best developed each bearing one pair of appendages (pereiopods).

The eleventh pair of appendages, in females, are modified into a round capsule to carry eggs. The pleon is composed of a variable number of segments (from 5 to 14), all bearing spines and a pair of appendages (pleopods). The last abdominal segment, called telson, is provided with uropoda which can be very long (Fig.15).

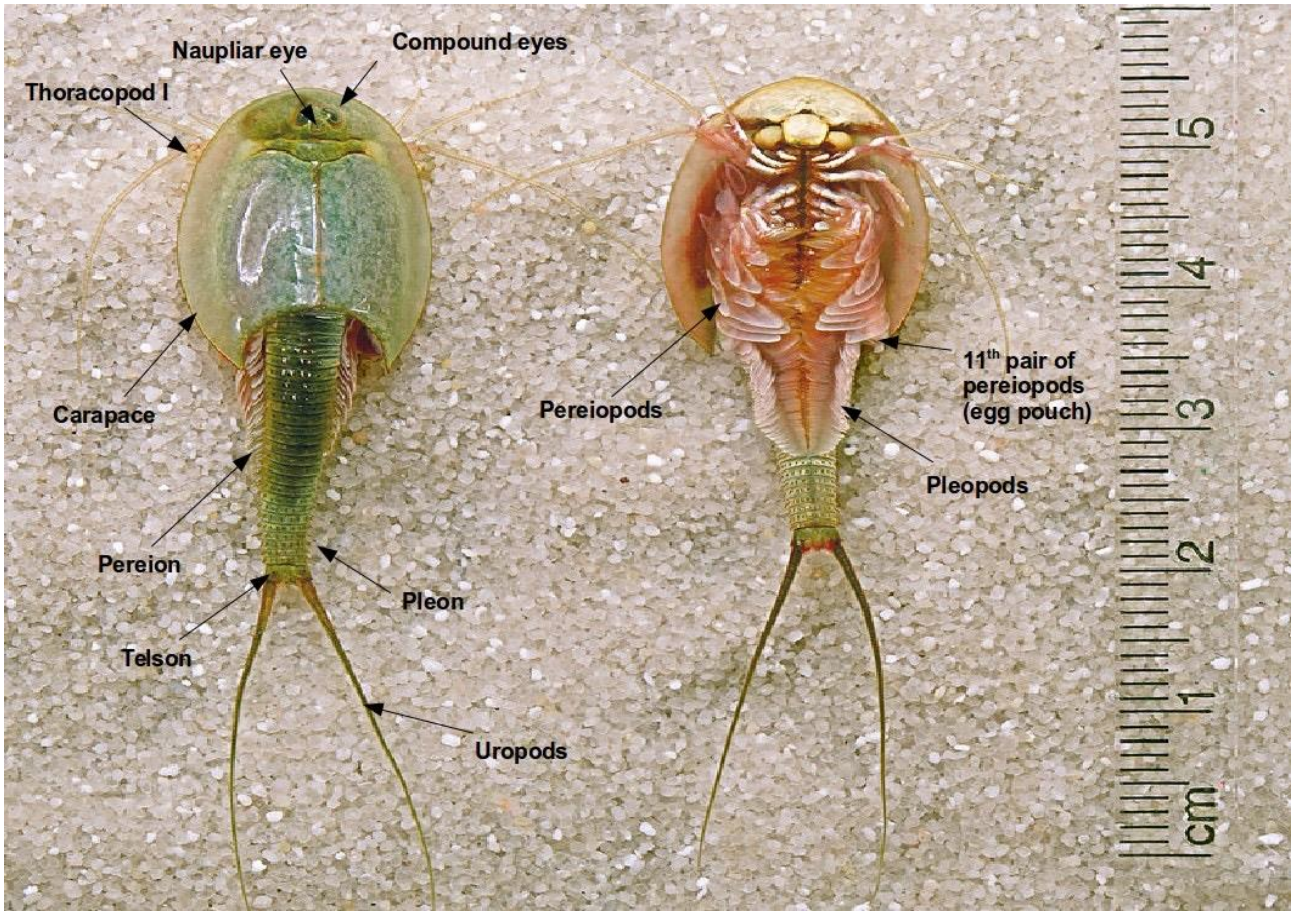


Fig. 15 Ventral and dorsal view of adult *Triops longicaudatus*. (Modified from

https://en.wikipedia.org/wiki/Triops_longicaudatus).

The morphological characters which allow the identification of the two genera are: the supra anal plate on the telson of *Lepidurus* species, lacking in *Triops* individuals, and the presence or absence of the second maxilla. The tadpole shrimps are morphologically indistinguishable from their ancestors in the Triassic period some 200 million years ago. For this high morphological stasis for a long evolution time, they are also known as 'living fossils'. 'Living fossil' is a term used, for the first time, by Charles Darwin in “On the Origin of Species” to describe species which were morphologically identical to their ancestors and with limited recent diversification, as if these organisms had stopped evolving. Examples are given by cycads, coelacanths, horseshoe crabs and *Ginkgo biloba.* However, notwithstanding their unchanged morphology, these species should adapt and evolve like others; in particular, recently, it was demonstrated that Notostraca were subject to two bursts of global radiation, one of them very recent, questioning the Darwin’s original concept of the term ‘living fossil’ (Mathers et al., 2013a).

Notostraca are present in all continents except Antarctica. They live in temporary aquatic systems that are most abundant in arid and semiarid areas. Nevertheless some *Lepidurus* species have been found in the artic area. Their ability to face these peculiar habitat is due to their diapausing eggs which survive even for several years and hatch when the environment condition become favorable. Their dispersal relies on abiotic factors (wind and water) or biotic vectors (birds, insects, amphibians and mammals, including anthropic activities). Their global distribution is especially due to their antiquity and their drought-resistance cysts (Vanschoenwinkel et al., 2012).

1.5.1. *Triops* genus Reproduction

The reproductive strategies implemented by Notostraca are various, including traditional gonochorism, parthenogenesis, hermaphroditism and androdioecy (coexistence of male and hermaphrodites). A single species, as *Triops cancriformis* Bosc, 1801 can have all these different reproduction.

It is to be noted that it is quite difficult to determine *T. cancriformis* sex and therefore reproductive strategy at the population level: the absence of external sexual characters and the eggs presence/absence as the unique observable character do not allow to distinguish between females and hermaphrodites nor they can be used to discriminate juveniles. Also chromosome analyses did not revealed the presence of sex chromosomes, even owing to the very small dimensions of Notostraca chromosomes (Marescalchi et al.,

2005). Only histological analyses of the gonads may firmly define the individual sex. In fact, as an example, in a population in which all individuals are provided with eggs, the histological presence of testicular lobes indicates a hermaphroditic reproduction. Another problem in sex population determination is due to the possible mixing of populations with different reproductive modes through resistant egg dissemination. For example, Scanabissi et al. (2005) found some functional males in an Austrian hermaphroditic population of *Triops cancriformis*.

In the literature, therefore sex ratio and histological analyses have been usually used to infer population reproductive mode. On these bases, Central and Northern European *T. cancriformis* populations are considered hermaphrodite/androdioecious (Zierold et al., 2009), Iberian populations appear gonochoric, while in the Italian Peninsula given that only all female populations have been found, reproduction should take place through parthenogenesis (Mantovani et al., 2004).

The reproduction mode can influence the genetic structure and diversity in the *Triops* genus (Velonà et al., 2009). Several studies showed a low genetic diversity, large inbreeding values (F_{IS}) and large population differentiation, within *Triops* populations; all these occurrences can be explained by possible founder events and by the reproduction mode of each population (Zierold et al., 2009; Velonà et al., 2009; Mantovani et al., 2008; Stoeckle et al., 2013).

During the 2000s, several research were conducted to investigate a possible link between reproductive modes and genetic diversity. Many molecular markers were tested for these genetic analyses: from several mitochondrial markers (such as cytochrome oxidase genes, mitochondrial control region, 12S and 16S) and different nuclear markers (28S and microsatellite loci). Most of these analyses agrees that, in general terms, *Triops* populations have a very low level of genetic variability (Cesari et al., 2004; Velonà et al., 2009; Zeirold et al., 2009; Horn et al.,2014). The study of Mantovani et al. (2008) highlighted also one possible sex-link locus in Italian and Spanish samples. Recently, the researches of Mathers et al. (2015), through the restriction site-associated DNA sequencing (RAD-seq) approach, have also identified sex-linked markers in the gonochoric and androdioecious European populations of *T. cancriformis.* They analyzed both males and females from a gonochoric population and males and hermaphrodites from an androdioecious population. For both populations they found a common set of sex-linked markers which could suggest the presence of preserved sex-determining region. Their analyses have also led to confirm a ZW sex determination system in *Triops cancriformis.* This system was found in two other androdioecious branchiopod species, *Eulimnadia*

texana and *Triops newberryi* (Sassaman, 1991; Sassaman and Weeks, 1993), thus suggesting that this sex determination mechanism is very conserved in Branchiopoda.

1.5.2. The difficult phylogeny of the order Notostraca

As written above, the “living fossils” of Notostraca order have a morphology indistinguishable from their ancestors of the Upper Triassic. However, in contrast with a highly conserved general morphology, notostracans exhibit a high plasticity regarding phenotypic characters within lineages making it difficult to determine the species boundaries, in addition to the absence of known morphological characters of taxonomic significance (Longhurst, 1955). For these reasons Notostraca taxonomy and phylogeny have been subjected to frequent revision. Linder in 1952 and Longhurst in 1955 and then Lynch in 1966 had reduced the number of species to only eleven worldwide. This taxonomy was accepted for many years. Only with the use of molecular techniques the need for taxonomic revisions due to cryptic diversification emerged, so that other species and subspecies were added.

Usually, the phylogenetic analyses were restricted geographically (Sassaman et al., 1997; Mantovani et al., 2008; Korn et al., 2010) or taxonomically (Korn et al., 2006) and were conducted with two or three DNA markers, resulting in poor resolution especially for analyses conducted at the intra-genus level (Vanschoenwinkel et al., 2012; Mantovani et al., 2004). However, most of these phylogenetic analyses had already highlighted the possible presence of cryptic species for both genera.

The latest works on Notostraca phylogeny were performed either using seven mitochondrial and nuclear markers on 38 extant Notostraca species (Mathers et al., 2013a) or using samples from all continents analyzed for four molecular markers (Korn et al., 2013). Both works confirmed that the split *Triops* - *Lepidurus* is placed in the Mesozoic. The calibration with a more parsimonious interpretation of the fossil record suggested an age of 152.3–233.5 Mya (95% high posterior density range). These analyses also supported the previously disputed monophyly of *Triops* and *Lepidurus* with good statistical support (ML bootstrap/ Bayesian posterior probabilities/MP bootstrap): 100/1.0/100 for *Lepidurus* and 96/1.0/100 for *Triops* node support (Fig.16; Korn et al., 2013). The presence of cryptic species was also highlighted. In fact, the results demonstrated that the diversity within *T. granarius* has been underestimated (Korn et al., 2013). *L. apus lubbocki,* as suggested before by Mantovani et al. (2009), was found to be a probable separate specie. Cryptic diversity was also identified in *L. couesii*: this species has a well known North American distribution, but it was recently found also in Apulia (South Italy) and

Sardinia. The latter lineages of *L. couesii* have a high level of divergence from American ones. This diversity led to propose them as new species (Mathers et al., 2013b). All these results reinforce the need for further future taxonomic revisions, in an attempt to further clarify the Notostraca phylogeny.

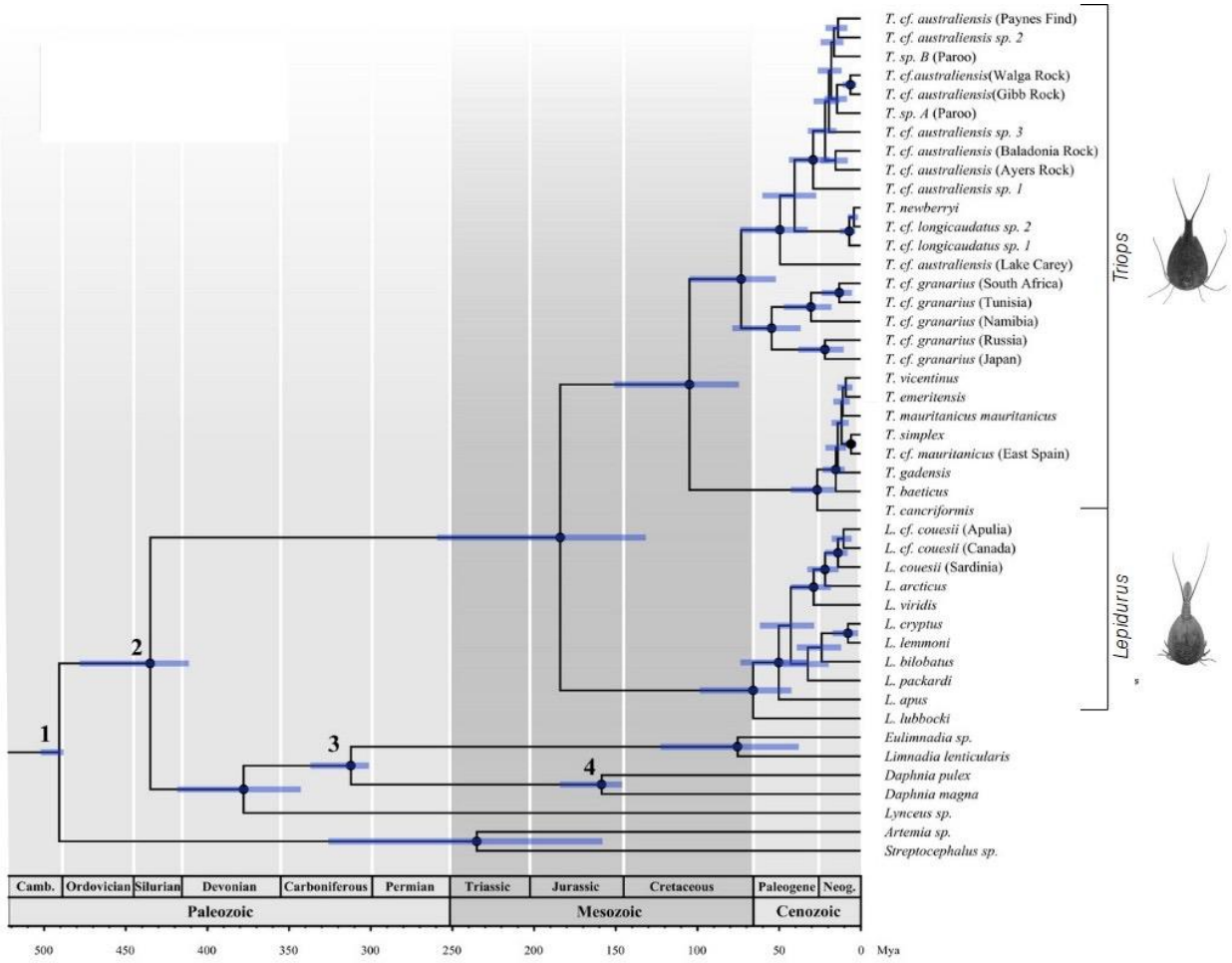


Fig. 16 Time calibrated phylogeny. Numbers at nodes correspond to the fossil calibrations: 1 *Rehbachiella*; 2 *Castracollis*; 3 *Ebullitiocaris elatus*; 4 *Daphnia* and *Ctenodaphnia* sp.. Black circles at the nodes indicated ML Bootstrap support values greater than 70 and posterior probabilities greater than 95. Error bars show the 95% confidence intervals of divergence times (modified from Mathers et al., 2013a).

1. RESEARCH AIMS

TEs are universal components of all living organisms. They were identified for the first time in 1945 by Barbara McClintock; since that time, several studies were performed to better understood the structure and the transposition mechanisms of these new genomic entities. However, even today, new elements and peculiar transposition mechanisms are continuously being discovered. Also the relationships between TEs and their host genomes are still under debate. Different hypotheses were proposed to explain TEs dynamics in host genome; one of these propose that host reproductive strategies can influence TEs evolutionary dynamics (Hickey, 1982). In fact, bisexual organisms, through homologous chromosomes recombination and reassortment during meiosis and amphimixis, can control the spread and proliferation of mobile elements, while unisexual organisms would experience an increase of these elements density due to the inability to eliminate them through exclusive mechanisms of sexual reproduction. In a long term, the accumulation and proliferation of TEs in unisexual organisms would be detrimental to the host fitness, leading to host lineage extinction. Therefore TEs would be expected to become inactive and/or domesticated or be eliminated by genome purging (Dolgin et al., 2006).

In order to evaluate these hypothesis, my main interests during the PhD research period were the isolation and molecular characterization of TEs in organisms with non-canonical reproductive strategies. I performed my analyses in two organisms: the stick insects of the *Bacillus* genus and in the tadpole shrimps *T. cancriformis*. In both instances reproductive strategies range from bisexual gonochoric reproduction, to unisexual parthenogenesis, making them an excellent model for the study and characterization of TEs, as it allows us to investigate the effects of reproduction on the biology of these elements.

2.1.Main results in *Bacillus* genus

The species of *Bacillus* genus show a great variability and plasticity in their reproductive strategies. The classical gonochoric bisexual reproduction seen in the majority of animal species is in the *Bacillus* genus present along with alternative reproductive strategies, such as parthenogenesis, hybridogenesis and androgenesis. In this genus, in particular, I focused on the R2 non-LTR retrotranposon.

R2 is one of the most investigated non-LTR retroelements that inserts into the ribosomal DNA locus. It is found in a variety of metazoan genomes, its wide distribution being possibly attributable to its specifically insertion into the 28S rRNA gene sequence.

On the whole, I isolated and sequenced by primer walking seven R2 complete elements: one in *B. rossius redtebacheri* (gonochoric population); one in *B. atticus atticus* (obligatory parthenogenetic population); two in *B. grandii grandii* (gonochoric population); one in *B. grandii benazzii* (gonochoric population) and two in *B. grandii maretimi* (gonochoric population).

The R2 element of *B. rossius redtebacheri* ($R2Br^{deg}$) appears degenerate, its ORF showing 14 frameshift mutations and one stop codon. In addition, analyses on the ability of the ribozyme to self-cleave the 28S/R2 co-transcript showed a very low ribozyme activity in $R2Br^{deg}$. One R2 element in *B. grandii grandii* ($R2Bg^{B}$) is the most divergent at the nucleotide level from the other *Bacillus* R2 elements, showing divergence ranging from 55.9% ($R2Bg^{A}$) to 57.1% ($R2Bm^{del}$). In addition it shows two Zinc Finger (ZF) motifs, at variance of the usual one ZF motif of all other *Bacillus* R2 elements; it probably represents a different R2 lineage. The two R2 elements found in *B. grandii maretimi* - R2Bm and $R2Bm^{del}$ - are nearly identical (0.4% of divergence) main difference being given by a 426 bp deletion in the latter. R2Bm shares with all other elements (R2Ba, $R2Bg^{A}$, R2Bg) the typical structure of a R2 element: length ranges from 2.7 to 4 kb, functional ORF, one ZF motif, RT and RLE domains and a 3' terminal poly-(A) tail with different nucleotide numbers. The nucleotide divergence analyses showed an unusually low divergence between R2s from *B. g. maretimi* and *B. g. benazzii* and the R2 functional element of the *B. rossius* congeneric species (3.8% - 5.2%). This observation led to think of a possible case of horizontal transfer (HT). All analyses that were performed to infer on HT – i.e. analyses of patchy distribution, absence of correlation between RT amino acid divergence and host divergence and incongruence between host and R2 phylogenetic trees – suggested possible HT events of the R2 element in the *Bacillus* genus.

These results are showed in detail in the following paragraphs.

2.1.1. Evolutionary dynamics of R2 retroelement in the genome of bisexual and parthenogenetic *Bacillus rossius* populations.

TEs dinamics in the host genome appears influenced by reproductive strategies. I started my research widening the analyses on *B. rossius* R2: I therefore managed in isolating/sequencing by primer walking an R2 complete element showing a degenerate ORF with 14 frameshift mutations and one stop codon in a gonochoric population of *B. rossius redtenbacheri* (Patti, Sicily). I then verified the distribution of this degenerated (R2Br^{deg}) R2 variant in comparison with the already known functional (R2Br^{fun}) element (Mingazzini, PhD Thesis, 2011) in thirteen Italian populations through PCR amplification of the 3' half of the ORF, including a diagnostic frameshift mutation. In particular I considered populations of both *B. rossius rossius* and *B. rossius redtenbacheri*, either gonochoric or parthenogenetic. The presence of R2Br^{deg} in *B. rossius tripolitanus* A indicates that this degenerate element was already present in the ancestral *B. rossius* genome since more 5 Myr ago, before the separation of Europe and North Africa that led to allopatric *B. rossius* lineage break (Mantovani et al., 2001).

Sequence variability of all elements was analysed for different aspects (variant distribution, selection acting, replication patterns and host phylogenetic history), all the analysis being not significant. On the whole, these results would indicate that no link exists between sequence variability and reproductive strategies in *B. rossius* subspecies.

This part of my work is presented in the paper:

Bonandin L., Scavariello C., Luchetti A. and Mantovani B. 2014 Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta, Phasmida). Insect Molecular Biology 23(6): 808-820.

The results were also presented at the following symposia/Congresses:

Claudia Scavariello, Livia Bonandin, Andrea Luchetti, Barbara Mantovani 2013. Non-LTR retrotransposon R2 molecular characterization and activity in *Bacillus rossius (*Phasmida, Bacillidae). V Congress of the Italian Society of Evolutionary Biology, SIBE

Oral communication pag. 57 (Atti di convegno-abstract). Trento, 28-31/08/2013.

Insect Molecular Biology

Insect Molecular Biology (2014) **23**(6), 808–820 doi: 10.1111/imb.12126

Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta Phasmida)

L. Bonandin*, C. Scavariello*, A. Luchetti and B. Mantovani

Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Bologna, Bologna, Italy

Abstract

Theoretical and empirical studies have shown differential management of transposable elements in organisms with different reproductive strategies. To investigate this issue, we analysed the R2 retroelement structure and variability in parthenogenetic and bisexual populations of *Bacillus rossius* stick insects, as well as insertions inheritance in the offspring of parthenogenetic isolates and of crosses. The *B. rossius* genome hosts a functional (R2Br^{fun}) and a degenerate (R2Br^{deg}) element, their presence correlating with neither reproductive strategies nor population distribution. The median-joining network method indicated that R2Br^{fun} duplicates through a multiple source model, while R2Br^{deg} is apparently still duplicating via a master gene model. Offspring analyses showed that unisexual and bisexual offspring have a similar number of R2Br-occupied sites. Multiple or recent shifts from gonochoric to parthenogenetic reproduction may explain the observed data. Moreover, insertion frequency spectra show that higher-frequency insertions in unisexual offspring significantly outnumber those in bisexual offspring. This suggests that unisexual offspring eliminate insertions with lower efficiency. A comparison with simulated insertion frequencies shows that inherited insertions in unisexual and bisexual offspring are significantly different from the expectation. On the whole, different mechanisms of R2 elimination in unisexual vs bisexual offspring and a complex interplay between recombination effectiveness, natural selection and time can explain the observed data.

Keywords: *Bacillus rossius* stick insects, bisexuality, parthenogenesis, R2 non-LTR element, retrotransposon insertion frequency, ribosomal DNA (rDNA).

Introduction

All living organisms contain in their genome a significant fraction of transposable elements (TEs), i.e. interspersed repeats able to move independently from one genomic location to another. They can be classified as class I TEs, moving via an RNA intermediate through 'copy-out/copy-in' (non-LTR elements) or 'copy out/paste in' (LTR elements) mechanisms, and class II elements, moving via a DNA intermediate by 'cut and paste' processes. TEs affect the host genome through a number of interactions that can be either beneficial or deleterious: they can modify gene expression and gene structure or promote recombination and genomic rearrangements (Kazazian, 2004).

Although they can be occasionally beneficial, TEs act as selfish elements and can replicate to high copy numbers affecting the host fitness (Doolittle & Sapienza, 1980). Mechanisms limiting TEs' proliferation, however, have evolved to allow host survival: cellular endogenous mechanisms, for example, silence TEs by methylation or through small RNA interference (O'Donnell & Burns, 2010). Nevertheless, much of the TEs population dynamics depend on the evolutionary dynamics of host organisms. In particular, theoretical studies linked differential TE loads with different reproductive strategies (Wright & Finnegan, 2001; Nuzhdin & Petrov, 2003 and reference therein). The idea that TEs can disperse in a population under bisexual reproduction was first put forward by Hickey (1982). Furthermore, according to Muller's ratchet

First published online 18 August 2014.

Correspondence: Dr Andrea Luchetti, Dip. BiGeA, via Selmi 3, 40126 Bologna, Italy. Tel.: + 39 051 2094173; e-mail: andrea.luchetti@unibo.it

*Equally contributing Authors.

theory, only bisexually reproducing organisms should be able to control and limit the deleterious proliferation of TEs thanks to mutation, recombination and the independent assortment of chromosomes. This would give natural selection more opportunities to operate (Arkhipova, 2005; Dolgin & Charlesworth, 2006; Loewe & Lamatsch, 2008). For the same reasons, in selfing or non-recombining genomes, TEs should either accumulate until host lineage extinction or be eliminated by genome purging (Wright & Finnegan, 2001; Nuzhdin & Petrov, 2003 and references therein). Actually, the few studies that have specifically addressed this issue showed contrasting patterns: while comparisons between selfing and outcrossing relatives in *Arabidopsis* and *Caenorhabditis* showed that TEs persist longer in selfing lineages, in obligate parthenogenetic lineages of *Daphnia pulex* the opposite result emerged (Valizadeh & Crease, 2008; Schaack *et al*., 2010). Moreover, in the wasp *Leptopilina claviceps,* only DNA transposons were found to have a higher copy number in the parthenogenetic lineage (Kraaijeveld *et al*., 2012).

Early-branching class I TEs show the particular feature of insertion into specific sequences, often tandem repeats; this strategy has been interpreted as an adaptation in order to bring little damage to the host genome, thus escaping genome purging (Malik *et al*., 1999; Malik & Eickbush, 2000). Among these TEs, R2 elements are the best studied retrotransposons so far and serve as a model to understand the biology of transposition of the whole non-LTR element subclass (Eickbush, 2002).

R2 is a non-LTR retroelement that inserts into the ribosomal DNA locus. R2 has a single open reading frame (ORF) flanked by two untranslated sequences of variable length. The ORF comprises the central reverse transcriptase (RT) domain, the DNA-binding motifs at the N-terminus and the restriction enzyme-like endonuclease (RLE) domain at the C-terminus. The C-terminal end of the R2 protein includes a cysteine-histidine (zinc finger, ZF) motif (CCHC), while the N terminal domain can contain one (CCHH), two (CCHH + CCHH or CCHC + CCHH), or three (CCHH + CCHC + CCHH) ZF motifs (Kojima & Fujiwara, 2005; Luchetti & Mantovani, 2013). R2 multiplies its copy number through a 'copy–out/copy-in' mechanism called target-primed reverse transcription (Eickbush, 2002; Eickbush & Eickbush, 2007). Target-primed reverse transcription frequently generates variable length deletions at the 5′ end of R2 and accounts for the numerous 5′ truncated copies in some insects (Burke *et al*., 1993). These 5′ truncations probably derive from the cellular degradation of the RNA transcript or from the inability of the reverse transcriptase to reach the 5′ end of the transcript (Christensen *et al*., 2006).

R2 is found in a variety of metazoan genomes and it is vertically inherited (Kojima & Fujiwara, 2005; Luchetti & Mantovani, 2013). The wide occurrence of R2 from Cnidaria to vertebrates is possibly attributable to the element niche provided by the ribosomal DNA locus (rDNA): in particular, R2 specifically inserts into the 28S rRNA gene sequence. This very specific target sequence makes the study of R2 activity quite interesting because of the interconnected TE-rDNA genomic dynamics.

rDNA shows a pattern of nucleotide variability known as concerted evolution, occurring when tandemly repeated sequences are more similar to each other within than between reproductive units (populations, subspecies, species, etc). This is achieved by a dual process: at the genome level, by molecular drive, through the homogenization of paralogous sequences by unequal DNA exchanges (mainly unequal crossing over, gene conversion, replication slippage; also known as genomic turnover mechanisms, GTM) and, at the reproductive unit level, by fixation of specific mutation profiles by means of bisexual reproduction (Eickbush & Eickbush, 2007; Plohl *et al*., 2008). The interplay between R2 dynamics and rDNA molecular drive is suggested by the deletion of rDNA units upon R2 insertion and rDNA unit replacement through compensatory GTM (Zhang *et al*., 2008; Mingazzini *et al*., 2011). This can be advantageous as 28S-inserted copies can be eliminated and replaced with new, functional 28S copies but, in this way, new sites for further R2 insertions are also provided (Eickbush & Eickbush, 2007; Zhang *et al*., 2008).

The short-term R2 inheritance and the interplay with rDNA have been well investigated in *Drosophila* spp. (Pérez-Gonzalez & Eickbush, 2001; Zhang & Eickbush, 2005; Zhang *et al*., 2008; Zhou & Eickbush, 2009; Zhou *et al*., 2013) and a few comparable studies address this issue in other organisms (Kagramanova *et al*., 2010; Ghesini *et al*., 2011; Mingazzini *et al*., 2011). The characterization of R2 elements in the genus *Bacillus* (Insecta Phasmida) should provide, however, the opportunity for analysis linking R2 dynamics to reproductive strategies. The genus is in fact a well-known example of reticulate evolution, and it includes the two highly differentiated species *Bacillus rossius* (bisexual with facultative parthenogenetic populations) and *Bacillus grandii* (strictly bisexual) as well as the obligatory parthenogenetic taxon *Bacillus atticus* and their related diploid and triploid non-Mendelian hybrids: the hybridogenetic lineages *B. rossius-grandii* and the obligatory parthenogenetic diploid *Bacillus whitei* and triploid *Bacillus lynceorum*. *B. rossius* is distributed over most of the Mediterranean basin with eight zymoraces, while *B. grandii* subspecies (*B. grandii grandii, B. grandii benazzii,* and *B. grandii maretimi*) and the hybrids are endemic to the Sicilian area; *B. atticus*, finally, occurs in the eastern part of the basin (Scali *et al*., 2003 and references therein; Fig. S1).

In the genus *Bacillus*, parthenogenesis occurs through a variety of mechanisms, ranging from automictic to apomictic processes. In *B. rossius*, in particular, females of parthenogenetic populations produce through meiosis haploid eggs; these start the development and then diploidize when reaching some thousands of cells, the mechanism generating an all-female offspring homozygous at all loci (Scali *et al*., 2003 and references therein).

Parthenogenesis may have major effects on concerted evolution: at variance of bisexual species, obligatory parthenogenetic taxa of the genus *Bacillus* show little or no fixation at all of either pericentromeric tandem repeats (Cesari *et al*., 2003; Luchetti *et al*., 2003) or rDNA intergenic spacer tandem repeats (Ricci *et al*., 2008). Generally speaking, it is to be recalled that a concerted evolution pattern in rDNA repeats can partly be fostered by selective pressures acting on both coding (18S, 5.8S and 28S genes) and regulatory (tandem repeats) regions (Ricci *et al*., 2008; Ambrose & Crease, 2011).

In the present study we report the characterization of R2 elements in the genome of *B. rossius* stick insects, along with inheritance studies in bisexual and parthenogenetic populations.

Results

R2 sequence characterization

The R2 element isolated from the *B. rossius rossius* population from Capalbio (roCAP) is 3492 base pairs (bp) long, excluding the 3′ terminal poly-(A) tail of about 23 nucleotides, with an ORF of 3165 bp encoding a protein of 1054 amino acids. The amino acid sequence shows an RT domain, an RLE domain at its C-terminal end and a single ZF motif (CCHH type) at the N-terminal end (Fig. 1). These observations are consistent with preliminary data from Mingazzini (2011). By contrast, the R2 element isolated from the *B. rossius redtenbacheri* sample from Patti (rePAT) is 3517 bp long and exhibits 14 frameshift mutations and one internal stop codon with respect to the roCAP R2 ORF (Fig. 1). The RT and the RLE domains as well as a single ZF motif at the N-terminal end are recognizable in the degenerate rePAT element. On the whole, the two R2 sequences are 9.2% divergent, the 5′ untranslated region (UTR) being more variable (16.4%) than either the ORF or the 3′ UTR (8.8% and 8.5%, respectively). For clarity, the R2 elements with a functional ORF and those with the degenerate one are hereafter referred to as R2Br^{fun} and R2Br^{deg}, respectively. The distribution of functional and degenerate R2 variants was studied in a population survey performed through PCR amplification and subsequent sequencing of the 3′ half of the element. Only the 1042 bp of the ORF, including the C-2480 frameshift mutation (Fig. 1, marked by the asterisk) were considered because, as a result of the presence of long poly-(A) tails, the sequencing of the 3′ UTR was not of sufficient quality. The three sampled populations of *B. rossius rossius* (roFOL, roCAP and roANZ) and the *B. rossius redtenbacheri* reBER sample show 100% functional elements; conversely, *B. rossius redtenbacheri* Sicilian samples (rePAT, reCDF, reMSN and reCUR) show 100% degenerated R2 sequences, while the remaining *B. rossius redtenbacheri* populations (reTDS, reVIR, reCOM and reGAB) and the *B. rossius tripolitanus* A population have both variants (Table 1, Fig. 2).

R2Br^{fun} sequence diversity within population ranges from 0.0012 ± 0.0007 to 0.0527 ± 0.0047. R2Br^{deg} variation is far more limited, being comprised between 0.0024 ± 0.0014 and 0.0162 ± 0.0057 (Table 1). On the

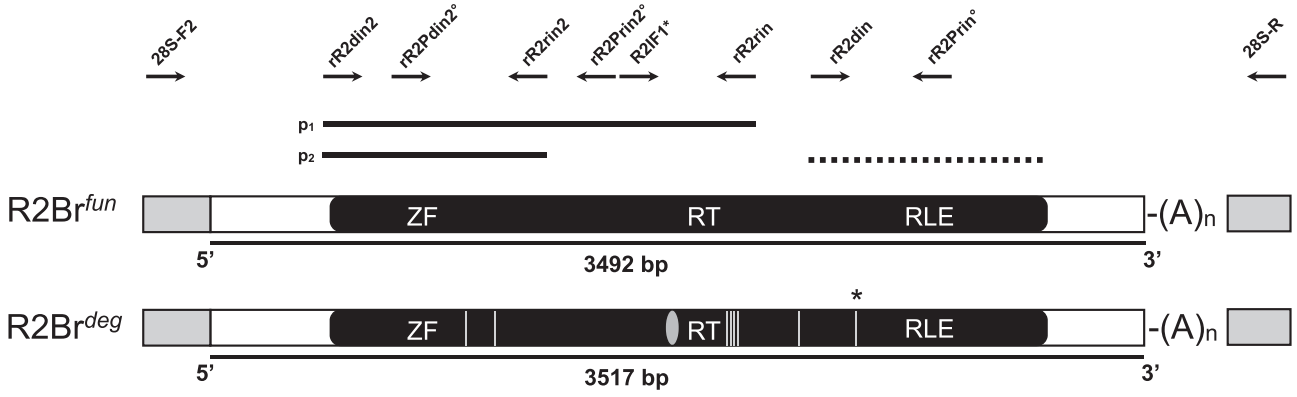


Figure 1. Schematic representation of R2Br functional (R2Br^{fun}) and degenerate (R2Br^{deg}) elements. Grey boxes indicate flanking 28S gene sequences; black boxes indicate the open reading frame (ORF) with the zinc finger (ZF), reverse transcriptase (RT) and restriction enzyme-like endonuclease (RLE) domains. In the R2Br^{deg} ORF, vertical grey lines indicate the frameshift mutations, the asterisk marking the C-2480 frameshift mutation used to distinguish the two R2 variants in the population survey; the oval represents the stop codon. In the panel above, the primers used for sequencing and transposon display are reported: the primer from Kojima & Fujiwara (2005) is marked with *, while those indicated with ° are primers specific for the degenerate variant. Thick bars p1 and p2 mark the regions used as probes in the transposon display blotting; the dashed bar indicates the region sequenced in the population survey.

Table 1. *Bacillus rossius* sampling with acronyms and reproductive strategy

Subspecies/collecting site	Acronym	RS	R2Br^{fun} (p-D ± SE)	R2Br^{deg} (p-D ± SE)
Bacillus rossius rossius				
Follonica	roFOL	P	7 (0.0075 ± 0.0025)	0
Capalbio	roCAP	G	10 (0.0032 ± 0.0011)	0
Anzio	roANZ	G	10 (0.0075 ± 0.0019)	0
B. rossius redtenbacheri				
Patti	rePAT	G	0	9 (0.0036 ± 0.0012)
Castanea delle Furie	reCDF	P	0	6 (0.0034 ± 0.0014)
Massa San Nicola	reMSN	P	0	6 (0.0054 ± 0.0018)
Curcuraci	reCUR	P	0	5 (0.0024 ± 0.0014)
Torino di Sangro	reTDS	G	12 (0.0028 ± 0.0011)	2 (0.0162 ± 0.0057)
Villa Rosa	reVIR	P	7 (0.0012 ± 0.0007)	1 (n/a)
Bertinoro	reBER	G	6 (0.0202 ± 0.0035)	0
Comacchio	reCOM	P	4 (0.0061 ± 0.0022)	5 (0.0038 ± 0.0016)
Gabonjin	reGAB	P	1 (n/a)	9 (0.0036 ± 0.0012)
B. rossius tripolitanus A				
Korbous	trKOR	G	8 (0.0527 ± 0.0047)	1 (n/a)

The number of functional (R2Br^{fun}) and degenerated (R2Br^{deg}) elements are given with their sequence diversity (p-D = p-distance) ± SE per population in parentheses. n/a, not applicable; RS, reproductive strategy; G, gonochoric; P, parthenogenetic.

whole, sequence variability between the functional and degenerate dataset was not significant (Student's *t*-test = 1.128, $P = 0.291$). Within the two datasets, even the comparison of R2Br variability between parthenogenetic and gonochoric populations was not significant ($t = 1.287$, $P = 0.263$; $t = 0.978$, $P = 0.505$).

The phylogenetic analysis included the two full-length elements and all 107 R2 3′ end sequences. These sequences split into two well-defined, main clusters: one includes only R2Br^{deg} elements, the other contains only R2Br^{fun} elements (Fig. 3A). Three sequences, though, fall outside the two main clades: one R2Br^{deg} from the reTDS population (reTDS-f1) and two R2Br^{fun} from the reGAB (reGAB-10) and trKOR (trKOR-9) samples (all indicated by arrowheads in Fig. 3A). The peculiar position of these sequences was explained by gene conversion events detected between functional and degenerate sequences: two functional elements showed extensive tracts exchanged with degenerate ones (309 bp in the reGAB-10 and 434 bp in trKOR-9), while only one R2Br^{deg} showed a converted tract (124 bp, reTDS-f1).

The median-joining network also confirmed the clear separation between functional and degenerate sequences, with the only exception of the trKOR-9

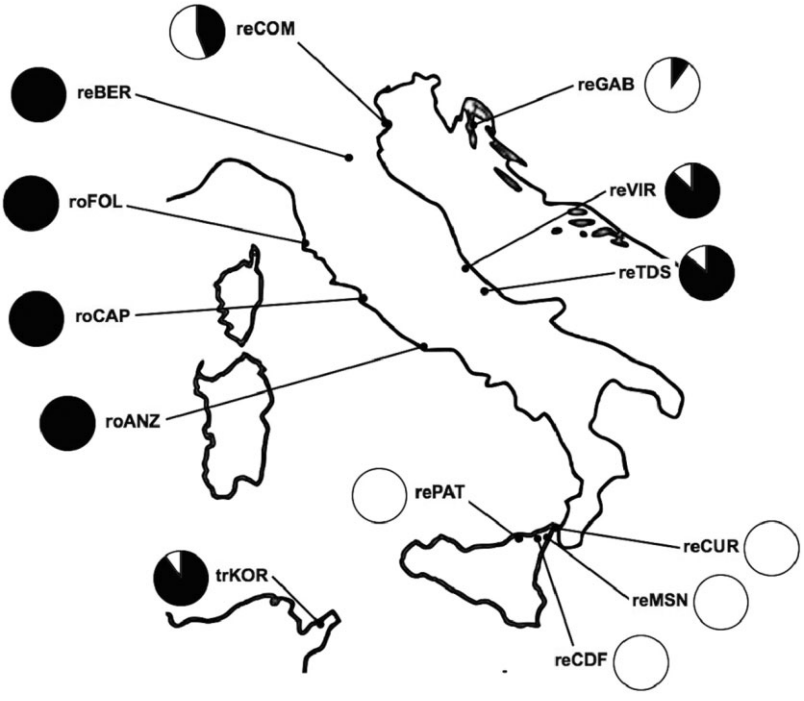


Figure 2. Geographic distribution of sampling localities. Pie charts indicate the proportion of R2Br functional (black) and degenerate (white) elements scored by sequencing in each sample. Acronyms are as in Table 1.

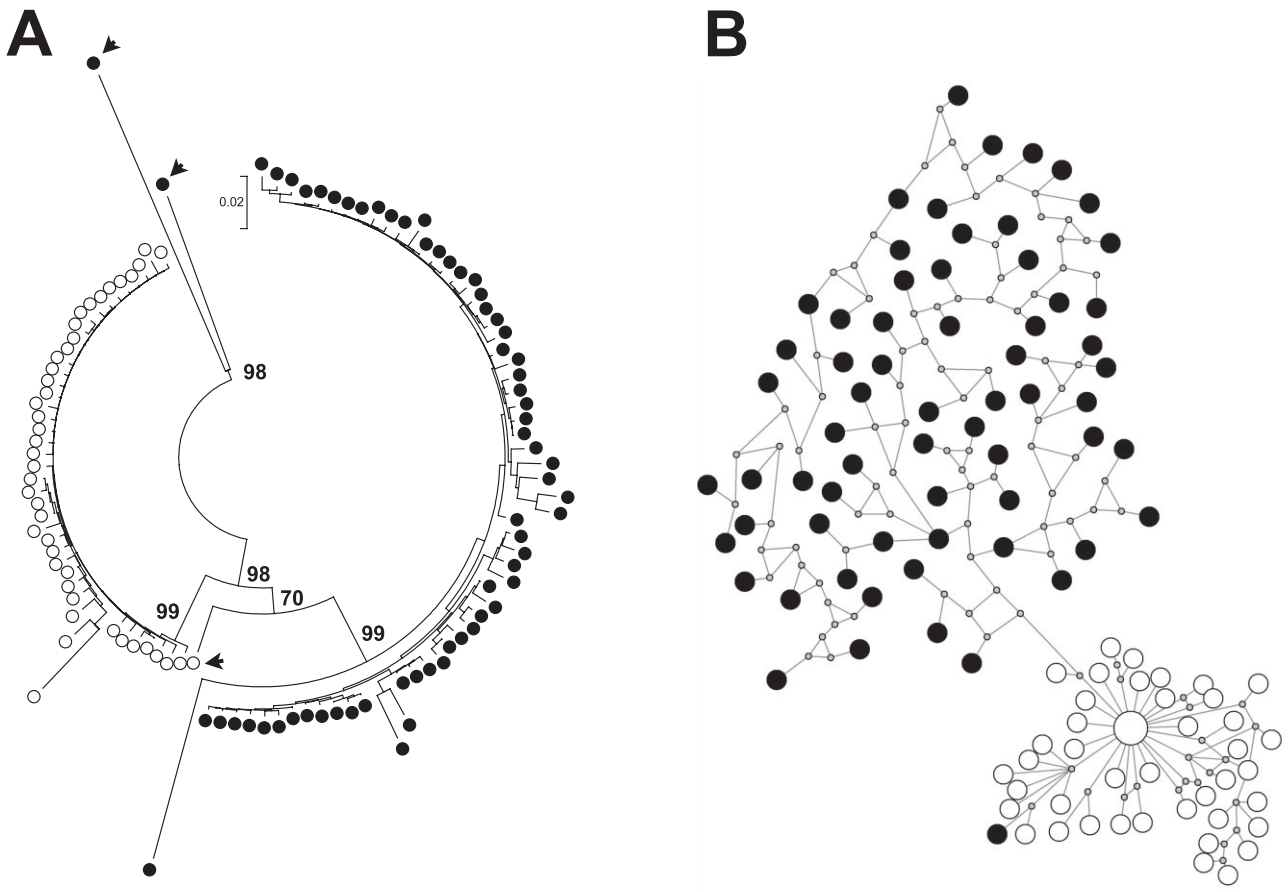


Figure 3. Phylogenetic analyses of R2 fragments sequenced in the population survey. (A) Minimum Evolution tree; numbers at nodes represent bootstrap values ≥70% obtained after 500 replicates. Arrowheads indicate sequences subject to gene conversion events. (B) Median-joining network; circles magnitude is proportional to the sequence variant frequency; the small, grey dots represent median-joining vectors. For both analyses, black-filled circles represent R2Br^{fun} sequences, while the white-filled ones indicate the R2Br^{deg} elements.

sequence. This sequence lacked the C-2480 frameshift mutation – and was therefore classified as functional – but otherwise showed a close relationship with R2Br^{deg} elements (Fig. 3B), owing to the above reported gene conversion event. The sub-networks corresponding to the two element types showed a quite different topology. The degenerate sequences exhibited a star-like pattern with a most common sequence, the ancestral one, from which all others originated. It is worth noting that the consensus R2Br^{deg} sequence is identical to this ancestral element. On the other hand, the sub-network built by functional sequences has a reticulated pattern, with no evident ancestry of a specific element (Fig. 3B).

Tajima's *D* values on R2Br sequences were significantly negative for both functional and degenerate datasets: $D = -2.685$, $P < 0.001$, and $D = -2.683$, $P < 0.001$, respectively. This indicates that both functional and degenerate elements are either under purifying selection or that they are both experiencing a relatively recent copy number expansion.

R2 insertions display

R2 activity was determined through the insertion display, the bands of variable length representing the 5′ end deletions occurring upon insertions (Pérez-Gonzalez & Eickbush, 2001). The detection of new 5′ truncated copies indicates insertion activity and, therefore, that the element is actively transposing (with a greater or lesser activity depending on the number of insertions). While this approach has some limitations, being unable to detect new insertions if the new truncated variant is of the same length as pre-existing bands or if the new insertion involves a full-length element (i.e. no. 5′ deletions occur), it has proven to be highly informative about the element's activity (Pérez-Gonzalez & Eickbush, 2001; Zhang & Eickbush, 2005; Zhang *et al.*, 2008; Zhou & Eickbush, 2009; Mingazzini *et al.*, 2011).

In parental parthenogenetic females, the number of sites occupied by R2 (S_p ♀) was in the range of 4–15 (Table 2). In their unisexual offspring, the total number of sites

Table 2. Parameters calculated on the basis of the R2 insertion display.

Parental individual(s)	S_p ♀	S_p ♂	S_p^{share}	S_p^{tot}	N_o	S_o^{tot}	S_o^{new}	ñ
reCDF-♀3	11			11	10♀	16	5	12.10
reCDF-♀4	14			14	10♀	18	4	12.80
reMSN-♀1	15			15	10♀	21	6	14.60
reMSN-♀2	4			4	10♀	4	1	3.60
reCUR-♀1	13			13	10♀	13	0	9.10
reCUR-♀2	15			15	10♀	17	2	12.10
reCUR-♀5 × roANZ-♂15	2	5	4	11	20♂	11	0	7.25
reCUR-♀6 × roANZ-♂9	0	0	4	4	20♂	4	0	3.70
roANZ-♀21 × roANZ-♂8	11	2	6	19	20♂	19	1	11.05
roCAP-♀1 × roCAP-♂4	2	8	5	15	10♂	19	4	9.50
roCAP-♀2 × roCAP-♂3	4	3	5	12	10♂	20	8	9.10

The number of sites occupied by R2 detected in parental individual(s), either in the single specimen (S_p♀, S_p♂) or shared between the two mates (S_p^{share}), and their total number ($S_p^{tot} = S_p♀ + S_p♂ + S_p^{share}$). The number of analyzed offspring individuals (N_o) and the total number of sites occupied by R2 in the offspring (S_o^{tot}) and the number of new sites (i.e. not occurring in parents) occupied by R2 (S_o^{new}); the estimated R2 copy number per haploid genome is indicated by ñ.

occupied by R2 (S_o^{tot}) was in the range of 4–21, with 1–6 new insertions (S_o^{new}) detected in all isolates with the exception of reCUR-♀1 progeny (Table 2). Interestingly, new sites occupied by R2 have been found also in those isolates that, from the sequence survey, showed only the presence of R2Br^{deg} (i.e. reCUR, reMSN, reCDF; Table 2), suggesting that these genomes could harbour active R2 copies.

In crossing experiments, the number of sites occupied by R2 was in the range of 2–11 within the female (S_p ♀) or the male (S_p ♂) parent, while 4–6 additional occupied sites are shared between the two mates (S_p^{share}; Table 2). The bisexual offspring showed S_o^{tot} in the range of 4–20; no new insertions were detected in the offspring of the two crosses involving former-unisexual females (reCUR-♀5 × roANZ-♂15 and reCUR-♀6 × roANZ-♂9), while 1–8 S_o^{new} were found in the descendants of the remaining three crosses (Table 2).

The R2 copy number per haploid genome (ñ) scored in the offspring varied widely: 3.60–14.60 in the unisexual dataset and 3.70–11.05 in the bisexual one (Table 2). On the whole, unisexual offspring showed a higher number of sites occupied by R2 (S_o^{tot}) than the bisexual offspring, 89 vs 73, although this difference was not significant ($t_{one\text{-}tailed} = 0.059$, $P = 0.477$). Moreover, the number of new inserted sites in the offspring (S_o^{new}) appeared higher in unisexuals than in bisexuals, but the difference was again not significant ($t_{one\text{-}tailed} = 0.220$, $P = 0.416$). The same also applies to the average R2 copy number per haploid genome in the unisexual–bisexual comparison ($t_{one\text{-}tailed} = 1.277$, $P = 0.117$).

By contrast, insertion frequency spectra, i.e. the frequencies' distribution of R2 insertions in the sample, were significantly different between unisexual and bisexual offspring (two-sample Kolmogorov–Smirnov test: $D = 0.351$, $P < 0.0001$; Fig. 4): in particular, high-frequency insertions (i.e. insertions occurring with a frequency >0.5) observed in unisexual offspring outnumber those found in bisexual offspring and most of the insertions remains close to 1.0 frequency. Conversely,

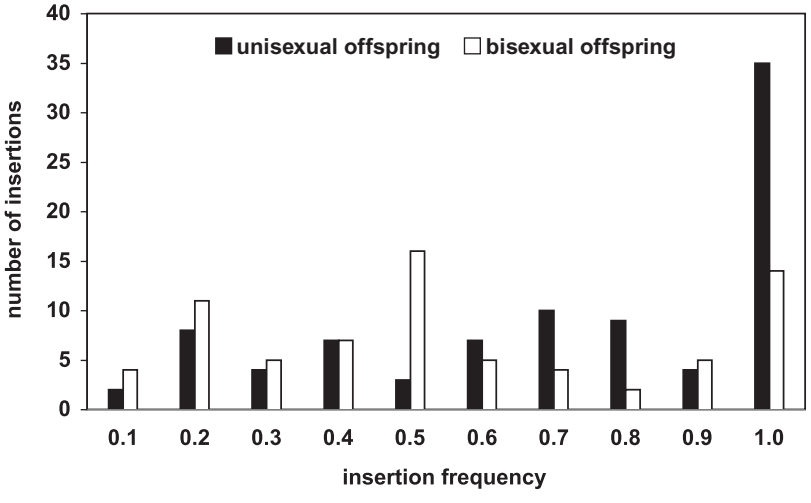


Figure 4. R2 insertion frequency spectra relative to unisexual and bisexual offspring. Each bin represents a class of insertion frequency in the offspring (i.e. the fraction of individuals carrying a certain R2 insertion over the total number of individuals) of either unisexuals (black) or bisexuals (white). On the *y*-axis are reported the number of R2 insertions recorded for each class.

low-frequency insertions (≤0.5) are more represented in bisexuals.

To obtain information about the contribution of recombination and selection in the elimination of R2 insertions, we derived expected data for unisexual and bisexual offspring: every deviation from strict insertion inheritance (in parthenogenetic offspring) or from Mendelian ratios (in crosses offspring) can be attributed to recombination or selection.

Based on the strict inheritance expected from the mechanism of parthenogenesis producing an all-homozygous progeny, all insertions of parthenogenetic mothers should have been present in the offspring at frequency = 1.0. However, data from unisexual progeny indicate quite a different scenario: excluding the new insertions occurring only in the progeny (S_o^{new}), maternally inherited insertion frequencies ranged from 0.1 to 1.0. Moreover, no offspring of any isolated female showed all mother's insertions occurring with frequency = 1.0 (not shown).

Expected insertion frequencies in crosses are, however, more difficult to establish with certainty as no information about the insertion heterozygosity can be derived (except for the two formerly unisexual females from the reCUR population that are homozygous). We thus simulated three different conditions: i) both parents are completely homozygous (all.ho); ii) both parents, with the exception of the reCUR females, are completely heterozygous (all.he); or iii) one parent is homozygous and the other one is heterozygous (one.ho). In all three simulations, bisexual offspring showed significant deviation from the Mendelian expectations of insertion frequencies, with the only exception of roCAP-♀1 × roCAP-♂4 descendants which fit the all.he condition (Table S1).

Observed R2 copy number per haploid genome is significant reduced from the expectation in unisexual offspring ($t_{paired} = 5.258$, $P < 0.01$; Fig. 5A). The same holds for bisexual offspring under the all.ho simulation ($t_{paired} = 3.630$, $P < 0.05$; Fig. 5B) but they did not show any significant difference from the expected inherited copy number under the all.he ($t_{paired} = 2.491$, $P = 0.067$) or one.ho simulation ($t_{paired} = 2.691$, $P = 0.055$), the R2 copy number resulting only slightly lower (Fig. 5B).

Discussion

R2Br^{fun} and R2Br^{deg} diversity and evolution

The present study first demonstrates the presence in the genome of *B. rossius* of two variants of the same R2 element: a functional one (R2Br^{fun}) and a degenerate one (R2Br^{deg}). The population sequence survey shows a peculiar distribution of the two variants in the Italian subspecies. On one hand, all *B. r. rossius* samples appear to host only R2Br^{fun}, and on the other hand, *B. r. redtenbacheri*

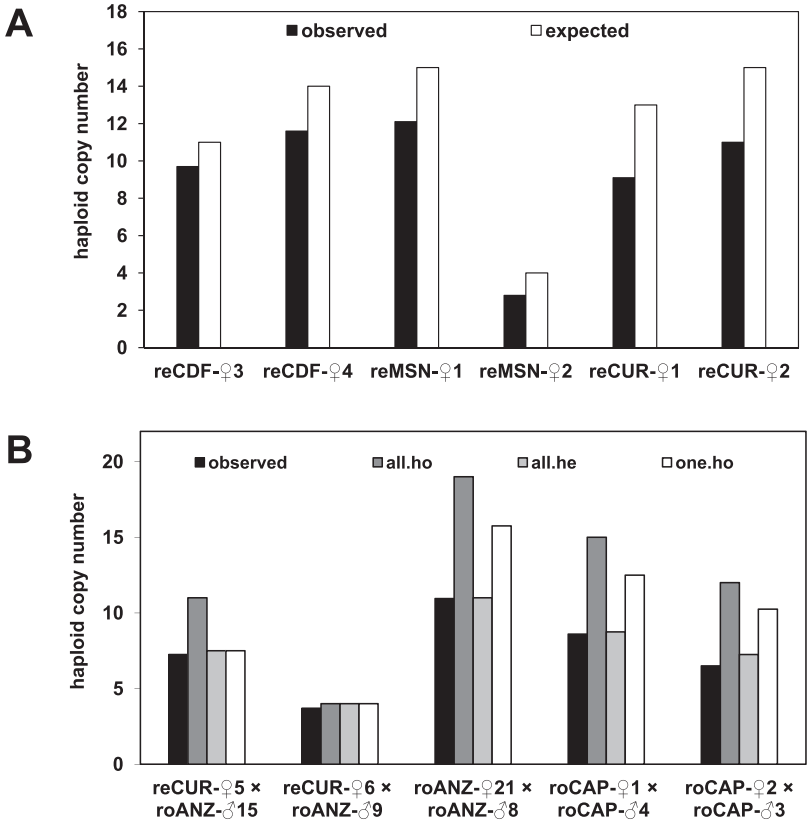


Figure 5. Observed and expected inherited R2 copy number in unisexual (A) and bisexual (B) offspring. Each bin represents parthenogenetic (A) or crosses (B) progeny observed (black) or expected (white and grey) R2 copy number per haploid genome, calculated as the sum of expected insertion frequencies based on simulation data. New insertions (i.e. those occurring only in the progeny but not in the parental individuals) are excluded. For unisexual offspring, the expectation is based on the fact that, given the parthenogenesis mechanism, progeny should have all the insertions of the mother. For bisexual offspring, as it is not possible to know if parental individuals have homozygous or heterozygous condition for the scored insertions, three simulations have been carried out: i) parental individuals are both homozygous (all.ho); ii) parental individuals, except reCUR females, are both heterozygous (all.he); iii) only one of the two parental individuals is homozygous while the other is heterozygous (one.ho). Acronyms are as in Table 1.

populations show widely different combinations: in all Sicilian populations only R2Br^{deg} was found, while peninsular populations contained both variants, in different percentages, or even only the R2Br^{fun} variant (reBER). The R2 insertion display, however, indicates that in the parthenogenetic *B. r. redtenbacheri* Sicilian populations reCUR, reMSN and reCDF new insertions occur, suggesting that some R2Br^{fun} copies could exist in their genome.

On the whole, no link appears between R2Br variant distribution and subspecies or reproductive strategies. Taking into account the evolutionary history of *B. rossius*, the presence of R2Br^{deg} in *B. r. tripolitanus* A indicates that it was already present in the ancestral *B. rossius* genome. When Europe and North Africa separated, leading to allopatric *B. rossius* lineage break (>5 Myr ago; Mantovani *et al*., 2001), R2Br^{deg} has been conserved in the extant genomes along with R2Br^{fun}.

Interestingly, the analysis of mutation distribution among sequenced R2Br ORF 3′ ends showed a non-random pattern. Tajima's *D* values calculated on both R2Br^{fun} and R2Br^{deg} datasets were significantly negative, suggesting that both variants are under purifying selection or in a state of a relatively recent copy number expansion. As purifying selection can be reasonably ruled out when dealing with mobile genetic elements, and particularly for the degenerate elements, Tajima's *D*s would indicate a recent burst of duplication. While this is perfectly conceivable for R2Br^{fun} and reflects its activity, the same cannot hold for R2Br^{deg} elements that do not code for a functional protein. Moreover, the median-joining network indicates that the replication of the two variants follows different patterns. Functional elements appear to duplicate through a multiple source model, reflected by the reticulated network, while degenerate elements seem to replicate via a master gene model, producing a star-like network. This implies that the R2Br^{deg} element differentiated once in the ancestral *B. rossius* genome and then gave origin to all extant R2Br^{deg} copies.

On the whole, therefore, data indicate that both R2Br variants can replicate, but their recent history shows the signature of different evolutionary dynamics.

While the sequence analyses provide evidence that R2Br^{deg} is undergoing duplication, the question remains how. At present we can only speculate about two possible explanations: i) R2Br^{deg}, after having lost its ability to code for a functional protein, became a non-autonomous element exploiting the retrotransposition machinery of an R2Br^{fun} not identified in our survey or ii) it duplicates when the host 28S sequence itself duplicates thanks to unequal crossing over or other GTMs. Both possibilities are likely to explain the observed data. In the *Drosophila* genome, non-autonomous elements derived from R2, or from a fusion between R2 and R1, (SIDE elements) were found to insert within the 28S by parasitizing the enzymatic machinery of parental elements (Eickbush & Eickbush, 2012). Moreover, again in the *Drosophila* genome, some R2 insertions have been found in multiple copies as a result of rDNA unequal crossing over, even if at a very low frequency (Zhou *et al*., 2013). Future studies addressing this specific issue will shed light on the mechanisms allowing R2Br^{deg} element duplication and, therefore, its survival.

R2Br insertion display and host reproductive strategies

The study of R2Br activity through the insertion display method, even if analysing a single generation, highlighted some interesting features. First of all, unisexual and bisexual offspring showed a similar number of R2-occupied sites (S_o^{tot}); however, unisexual offspring exhibited more high-frequency insertions than bisexual offspring, indicating that each insertion is less likely eliminated in the unisexual than in a bisexual condition. This pattern is partly consistent with theoretical predictions and with previous analyses of TE insertions in outcrossing and selfing lineages, where outcrossers showed more insertions at low frequency than the selfing relatives in both *Arabidopsis* spp. (Lockton & Gaut, 2010 and references therein) and *Caenorhabditis* spp. (Dolgin *et al*., 2008). These data contrast, however, with the dynamics of r*Pokey*, a rDNA-targeting element in the *D. pulex* genome. In this instance, obligate unisexuals have a significantly lower number of insertions with respect to cyclically unisexual isolates, but their frequency spectra are very similar (Valizadeh & Crease, 2008). This has been explained by the polyphyletic, recent origin of obligate parthenogenetic isolates that inherited the insertion profile of their cyclical ancestors and, then, possibly underwent clonal selection (Valizadeh & Crease, 2008). *B. rossius* bisexual populations can rapidly shift to unisexuality through male loss, as it has been directly observed in the wild (Scali *et al*., 2003), while the chance of sex chromosomes meiotic missegregation accounts for the appearance of males in laboratory-reared parthenogenetic lines (Mantovani B., pers. obs.). Multiple or recent shifts may explain why unisexuals and bisexuals have the same number of R2-occupied sites, but differences remain with *D. pulex*. One possibility is that the crustacean may undergo heavier environmental stresses than stick insects and this would cause either stronger clonal selection or, simply, stronger selection favouring genomes with a lower r*Pokey* load. This might also explain differences in relative frequency spectra, even though the type of parthenogenesis can play a role.

D. pulex eggs are diploid by abortive meiosis (Hiruta *et al*., 2010), therefore the heterozygosity of the mother is maintained in the offspring even if loss of heterozygous mutations has been documented by ameiotic recombination (Omilian *et al*., 2006). In contrast, *B. rossius* parthe-

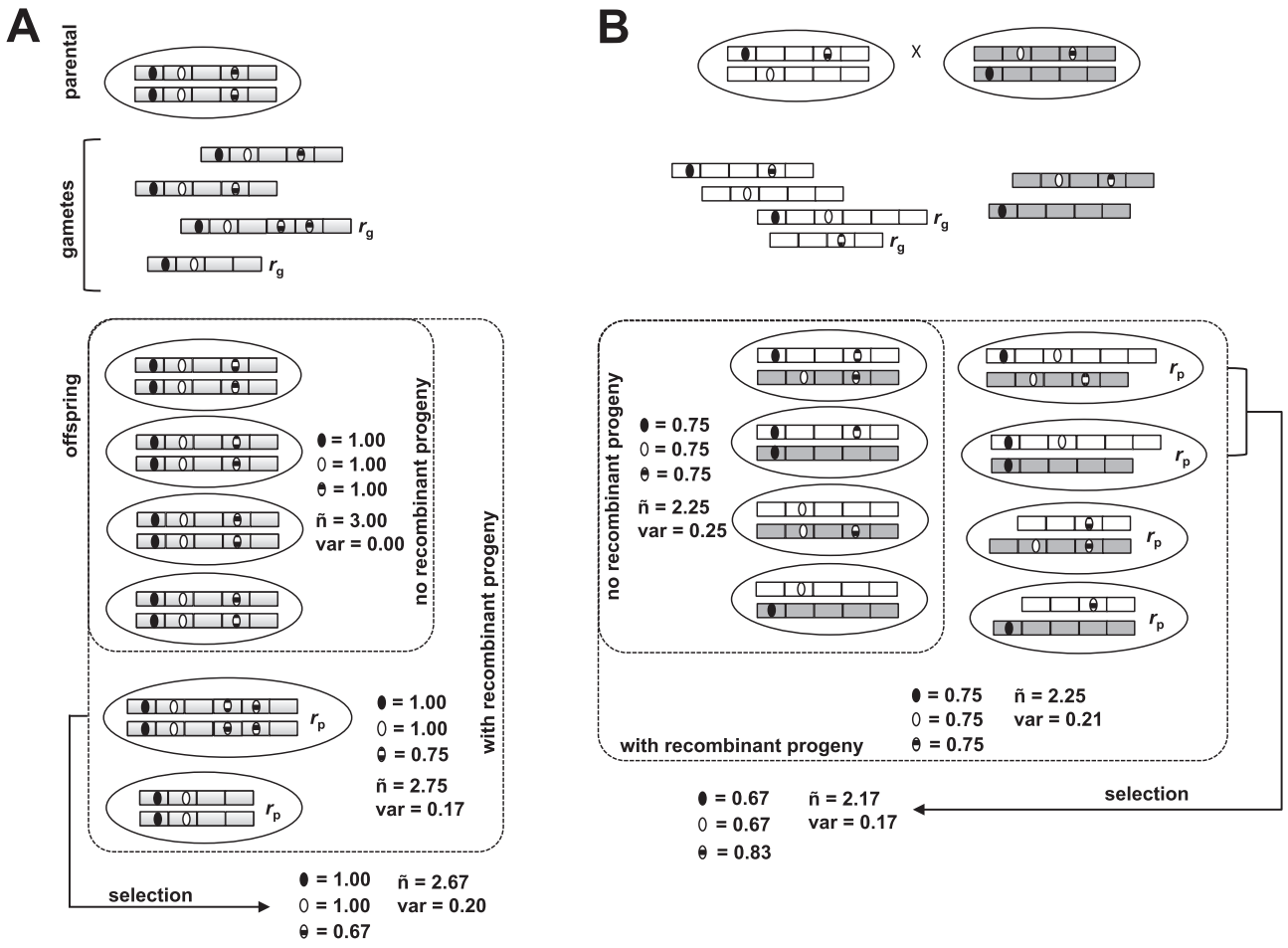


Figure 6. Scenarios of R2 insertion inheritance under unisexuality (A) and bisexuality, all.he simulation (B). Each box represents a 28S rRNA and dots are R2 insertions (different colours = different insertions); for each parental individual, gametes (r_g: recombinant gamete) and offspring (r_p: recombinant progeny) are reported. For three possible situations (progeny without recombinant; progeny with recombinant; progeny after selection) R2 insertion frequencies, copy number (ñ) and its variance (var) have been calculated. In this model, selection has been considered as acting on offspring with longer rDNA array. In the bisexual post-selection scenario, insertion frequencies are significantly different from those calculated under the two other scenarios ($G = 8.1$; d.f. = 3; $p < 0.05$).

nogenetic development is based on eggs deriving from a normal meiosis, therefore they start haploid and then diploidize. This mechanism generates a fully homozygous offspring that should inherit the 100% of the mother's insertions.

The comparison between presently obtained data and simulated data, however, clearly showed that observed frequencies of inherited insertions are significantly different from expected ones. In Fig. 6 and Fig. S2, we show possible scenarios that may explain the results of our simulations. Deviations from an insertion frequency of 1.0 in the unisexual progeny may only happen if unequal recombination (probably the most common GTM), possibly coupled with selection on particular genotypes, has eliminated some of the insertions (Fig. 6A). In our dataset, such recombination events should have occurred several times, as multiple insertion loss can be observed in all parthenogenetic offspring; however, from a parental homozygous condition, the probabilities of generating insertion-purged gametes are lower than if starting from a heterozygous condition. In fact, the insertion frequency distribution of unisexual is skewed toward high frequencies.

In crosses' progeny, with only one exception, Mendelian ratios are rejected in all simulated expectations and, quite interestingly, the observed inherited R2 copy numbers per haploid genome are slightly lower but not significantly different from the expected ones in the all.he and the one.ho simulations. If parents are all homozygous for R2 insertions (all.ho), the scenario does not change much from the unisexual condition, with the offspring expected to inherit all insertions at frequency =1.0. If parents are heterozygous (all.he simulation), recombination events, such as unequal crossing over, would simply reshuffle the rDNA, relocating the insertions within the array; thus, when eventually checked using transposon display, the observed frequencies and copy number would not be

changed in a single generation (Fig. 6B). Other unequal DNA exchanges such as gene conversions, actually, would probably lower the copy number in a single generation. On the other hand, selection on particular genotypes would modify the frequencies, although the copy number decrease is decidedly less than in unisexual offspring (Fig 6B), in line with our comparison between observed and simulated data. If both parents produce recombinant gametes, insertion frequency and copy number variation behave similarly (Fig. S2A). Generally speaking, however, whether or not recombination and/or selection occur, the bisexual progeny will show a greater copy number variance than unisexual offspring (Fig. 6B; Fig. S2A).

Under the one.ho simulated scenario, however, the management of R2 insertion seems to change substantially: unequal recombination is effective in one generation only if occurring in the homozygous parent, regardless of whether the heterozygous parent produces recombinants or not. Moreover, selection appears to have a weak effect on both deviation of insertion frequencies and copy number reduction (Fig. S2B–D). This is probably attributable to a buffering effect of the homozygous parent that will redistribute all its insertions to the offspring. It is likely that, in this instance, the process of R2 insertion elimination would be slower and, therefore, more evident in the subsequent generations.

Following the above explanation (Fig. 6; Fig. S2), the observed pattern, compared with simulated data, suggests two different mechanisms of R2 elimination in unisexual vs bisexual offspring. In the case of unisexual offspring, unequal crossing over generates two new genotypes, one with an insertion loss and one with insertion duplication. The genotype subject to the loss would give rise to a 'purged' clonal line, counteracting, theoretically speaking, the Muller's ratchet effect. The genotype with the duplication may be eliminated or not, depending on the effects on the fitness. It is to be noted, however, that such duplications would not be detected using the insertion display method, as duplicated insertions will unavoidably result in the same PCR amplified band. The scenario depicted in the one.ho simulation (Fig. S2B–D) supports this pattern, unequal crossing over having major effects when occurring in the homozygous parent.

In all-heterozygous bisexual offspring, the R2 insertion elimination would follow a more complicated route. Recombination does not seem to affect the retrotransposon load in a single generation. More importantly, outcrossing generates higher R2 copy number variance in the progeny, resulting in a higher variance of inserted/uninserted 28S rRNAs. This would give natural selection more options to operate in the following generations (Mingazzini *et al*., 2011); in fact, in this instance, selection has a greater effect than recombination.

Generally speaking, based on theoretical predictions such as Muller's ratchet theory, the retention of high frequency TE insertions in unisexual, selfing or asexual lineages is probably attributable to the absence (or very low efficiency) of truly effective recombination mechanisms. The data in the present study suggest that, whether recombination is effective or not, the time available (in terms of number of generations) for insertion elimination may also play a role. In situations where organisms are completely/highly homozygous or have very high selfing rates, more generations may be necessary to achieve elimination efficiencies similar to those of more heterozygous organisms. In line with this, variance in TE copy number should play a greater role in the fate of the whole TE load. Moreover, the rate of TE elimination in homozygous organisms is further contrasted by the TE insertional activity itself, continuously providing new TE copies.

Another interesting result of our analysis is that unisexual and bisexual offspring showed the same number of new R2 insertions (i.e. those occurring in the offspring only); this would indicate that R2Br has the same activity despite host's reproductive strategies, even if insertions within unisexual genomes persist longer than in the bisexual ones. It appears, therefore, as if no mechanisms are acting on R2 transposition rate in these parthenogenetic genomes. Suppression of transposition is a common phenomenon, mainly performed by micro-RNAs or methylation, aiming to keep at bay TEs activity in order to avoid potentially deleterious replicative burst (O'Donnell & Burns, 2010). This is particularly important in a homozygous genome such as that of parthenogenetic stick insects, because natural selection would have fewer opportunities to eliminate deleterious insertions. However, things can be slightly different for rDNA-targeting TEs: the specific insertion into repeated genes (as rDNA) undergoing GTMs will ensure that there would always be gene copies free from TE insertions, still allowing cellular functionality. Site-specificity might therefore be an adaptive strategy in order to provoke little damage to the host and then to escape purging (Malik & Eickbush, 2000). In fruit flies, however, R2 transcription and, therefore, transposition are largely controlled by heterochromatinization of densely inserted rDNA arrays followed by possible nucleolar dominance (Eickbush *et al*., 2008); this can be achieved by compartmentalizing functional R2s in restricted rDNA regions that can be transcriptionally silenced. Starting from a condition in which functional R2s are located within transcriptionally active rDNA copies, silencing can be obtained by either i) rearranging the array by GTMs (mainly unequal crossing over) so that R2s are moved to heterochromatinizable regions, or ii) the establishment of nucleolar dominance after mating with a partner bearing an R2-silenced rDNA array (Eickbush

et al., 2008). In the case of stick insect parthenogenetic populations, the latter does not appear possible, so that recombination would be a major factor maintaining R2 at bay. Based on the rate of R2 elimination evidenced by transposon display data, unequal recombination occurs at an appreciable rate as it can be deduced in isolates' offspring. This can be explained if we consider GTMs as part of a mechanism pushing R2 load toward regions of the rDNA array that can be silenced.

Experimental procedures

Sampling and DNA isolation

Specimens were field-collected from 13 localities (Table 1, Fig. 2) and either reared in plastic boxes to obtain offspring or frozen at –80 °C until molecular analysis. Total DNA was extracted from single stick insect legs or from the whole body of first instar larvae with the standard phenol/chloroform protocol.

R2 isolation and sequencing

R2 presence was first checked by a PCR assay, using a primer anchored on the 28S rRNA gene, downstream the R2 insertion site, and a collection of degenerate primers, complementary to the element ORF region (Fig. 1; Kojima & Fujiwara, 2005). PCR amplifications were performed in a 50-μl reaction mix using the GoTaq Flexi (Promega, Madison, WI, USA), following the manufacturer's instructions. Thermal cycling was as follows: initial denaturation step at 95 °C for 5 min, 35 cycles of denaturation at 95 °C for 30 s, annealing at 48 °C for 30 s and extension at 72 °C for 3 min, and a final extension at 72 °C for 7 min. PCR products were run on a 1.5% agarose electrophoresis gel and bands of ~1800 bp were eluted using the Wizard SV Gel and PCR Clean-Up System kit (Promega). Fragments were inserted into a pGEM-T Easy vector (Promega) and used to transform *Escherichia coli* DH5α cells. Recombinant colonies were PCR-amplified with T7/SP6 primers and sequenced at Macrogen Inc. – Europe Lab (Amsterdam, The Netherlands).

This approach led to the recovery of R2 3′ ends from which R2 complete sequences were obtained through primer walking coupling a 28S-anchored primer (28S-F2), annealing upstream of the insertion site, with specifically designed primers complementary to R2 internal regions (Table S2; Fig. 1). Full-length R2 elements were therefore isolated and characterized from one female specimen each of gonochoric populations of *B. rossius rossius* (Capalbio, Tuscany) and *B. rossius redtenbacheri* (Patti, Sicily; Table 1; Fig. 2). Partial R2 sequences at the 3′ end were then isolated from one individual for each of the 13 localities reported in Table 1 and Fig. 2 through PCR amplification using the primer pair rR2din/28S-R (Table S2; Fig. 1). PCR reaction conditions, cloning and sequencing were as described above.

R2 sequence analyses

Sequences were edited and assembled using MEGA v. 5.2 (Tamura *et al*., 2011) and ORFs were searched with the ORF Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Sequence alignments through the CLUSTALW algorithm, p-distance (p-D) ± SE and Minimum Evolution tree (with nodal support after 500 bootstrap replicates) were carried out using MEGA v. 5.2. Median-joining network (Bandelt *et al*., 1999) was calculated with Network v. 4.6 (Fluxus Engineering, Clare, Suffolk, UK). Tajima's *D* test for selection and the detection of gene conversion events were performed with DnaSP v. 5.1 (Librado & Rozas, 2009).

Estimation of R2 activity

The analysis was performed on selected parental individuals and in a sample (10–20 individuals) of their offspring (Table 2). In particular, we analysed the thelytokous offspring of two females for each of three parthenogenetic *B. rossius redtenbacheri* populations (Castanea delle Furie, Massa San Nicola and Curcuraci) and the male progeny of crosses involving as parental females either parthenogenetic specimens of *B. rossius redtenbacheri* from Curcuraci (reCUR ♀ X roANZ ♂, two crosses) or *B. rossius rossius* gonochoric females from Anzio (roANZ ♀ X roANZ ♂, one cross) or Capalbio (roCAP ♀ X roCAP ♂, two crosses). Males were chosen in the latter instances because they certainly represent descendants of bisexuality, while female progeny may still be of parthenogenetic origin. For each progeny, both early-hatched individuals and late-hatched individuals were chosen. This is of particular importance for cross descendants to be sure that they represent distinct meiotic products of the parental male.

The estimate of R2 activity was assayed through the insertion display technique: given that the R2 5′ end is subject to deletions of various lengths upon insertion, a single 5′ end deletion corresponds to a single insertion event. Through this technique, the various deletions occurring in a single individual can be visualized and, therefore, the number of insertions can be estimated. For R2 5′ end insertion display, we used a strategy based on PCR amplification, as both the element and the flanking region sequences are known. A 28S-anchored primer (28S-F2), annealing upstream of the insertion site, was alternately coupled with two primers, rR2rin2 and rR2rin, specifically designed to anneal 1241 bp and 1862 bp from the 5′ end, respectively (Table S2; Fig. 1). The two PCR reactions were performed with the GoTaq Flexi kit (Promega) following the manufacturer's instructions and the PCR conditions: initial denaturation at 95 °C for 2 min, 35 cycles of denaturation at 94 °C for 30 s, annealing at 48 °C for 30 s and extension at 72 °C for 1 min 30 s or 2 min (with the primers rR2rin2 and rR2rin, respectively) and a final extension at 72 °C for 5 min. PCR products were electrophoresed on a 1.5% agarose gel and the complete elements and every 5′ end deleted R2 resulted, thus, in a PCR amplification band.

The gel was then Southern blotted onto a positively charged nylon membrane. Membranes were hybridized with PCR-amplified R2-specific probes; the amplicon 28S-F2>rR2rin2 was hybridized with the probe rR2din2>rR2rin2, and the amplicon 28S-F2>rR2rin with the probe rR2din2>rR2rin, and the resulting signal was detected using the AlkPhos Direct Labelling and Detection System kit (GE Healthcare, Piscataway, NJ, USA). For each insertion display, two replicates were carried out to be sure of band scoring correctness. Data from the two PCR reactions were then combined to obtain a single 5′ end deletion profile per assayed individual and the number of bands visualized on the gel was calculated per individual and per offspring. These numbers are referred to as the number of sites occupied by R2 (S) and indicate the number of inserted 28S scored with this analysis. The following variables were calculated: the number of sites occupied

by R2 in parental individual(s) (S_p), the total number of sites occupied by R2 (S_o^{tot}) and the number of new sites occupied by R2 (new insertions; S_o^{new}) in the offspring.

Moreover, the R2 copy number per haploid genome ($\tilde{n}$) has been calculated in the offspring of both isolates and crosses as

$$\tilde{n} = \sum^{s} \frac{b_S}{N_O}$$

where b_S is the number of bands displayed at the S_{th} site occupied by R2 and N_O is the number of assayed offspring individuals (Charlesworth & Charlesworth, 1983). Insertion frequency spectra (i.e. the distributions of frequencies with which R2 insertions occur in the sample) were calculated, partitioning data on the basis of reproductive strategies: unisexual vs bisexual.

Expected frequencies in the progeny were calculated in the unisexual dataset taking into account that, because of their parthenogenesis mechanism, parental individuals are homozygous and so are their offspring (Scali *et al*., 2003): therefore, leaving aside new insertions, offspring should strictly inherit the same insertions as the mother. It is, thus, expected that the 100% of the mother's insertions should appear in the progeny (frequency = 1.0).

Expected frequencies in crosses' offspring, however, cannot be calculated so simply, as we have no information about the insertions' heterozygosity in parental individuals (except the two reCUR females that, being from parthenogenetic population, are homozygous); therefore, we drew three different simulations of Mendelian expectation for each R2 site: i) parental individuals are both homozygous (all.ho); ii) parental individuals, except reCUR females, are both heterozygous (all.he); and iii) only one of the two parental individuals is homozygous while the other is heterozygous (one.ho). In other words, we simulated simple Mendelian crosses between individuals considering their insertion pattern, revealed by the insertion display analysis, as being made by insertions all homozygous (all.ho), all heterozygous (all.he) or homozygous in one individual and heterozygous in the other one (one.ho). We, thus, obtained the genotype frequencies in the simulated progeny (from which we derived the expected insertion frequencies), assuming no recombination and no selection, and compared these data with those observed in the offspring insertion display experiments. The G-test was applied to check if observed insertion frequencies fit the expectation. Moreover, expected inherited R2 copy numbers per haploid genome based on expected insertion frequencies have been calculated and compared with the observed ones, computed as above, but excluding new insertions (i.e. S_o^{new}).

Data availability

Sequences have been submitted to GenBank under the following accession numbers: **KJ958674** ($R2Br^{fun}$), **KJ958675** ($R2Br^{deg}$) and **KJ958565-KJ958673** (partial R2 elements for population survey).

Acknowledgements

This work was supported by RFO and Canziani funding to B.M. We are grateful to Danna Eickbush for the critical reading of the manuscript and for fruitful discussions and to two anonymous referees which helped greatly in improving the manuscript.

References

Ambrose, C.D. and Crease, T.J. (2011) Evolution of the nuclear ribosomal DNA intergenic spacer in four species of the *Daphnia pulex* complex. *BMC Genet* **12**: 13.

Arkhipova, I.R. (2005) Mobile genetic elements and sexual reproduction. *Cytogenet Genome Res* **110**: 372–382.

Bandelt, H.J., Forster, P. and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**: 37–48.

Burke, W.D., Eickbush, D.G., Xiong, Y., Jakubczak, J. and Eickbush, T.H. (1993) Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. *Mol Biol Evol* **10**: 163–185.

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.

Charlesworth, B. and Charlesworth, D. (1983) The population dynamics of transposable elements. *Genet Res* **42**: 1–27.

Christensen, S.M., Ye, J. and Eickbush, T.H. (2006) RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc Natl Acad Sci USA* **103**: 17602–17607.

Dolgin, E.S. and Charlesworth, B. (2006) The fate of transposable elements in asexual populations. *Genetics* **174**: 817–827.

Dolgin, E.S., Charlesworth, B. and Cutter, A.D. (2008) Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis* nematodes. *Genet Res* **90**: 317–329.

Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**: 601–603.

Eickbush, D.G. and Eickbush, T.H. (2012) R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements. *Mobile DNA* **3**: 10.

Eickbush, D.G., Ye, J., Zhang, X., Burke, W.D. and Eickbush, T.H. (2008) Epigenetic regulation of retrotransposon within the nucleolus of *Drosophila*. *Mol Cell Biol* **28**: 6452–6461.

Eickbush, T.H. (2002) R2 and related site-specific non-LTR retrotransposons. In *Mobile DNA II* (Craig, N., Craigie, R., Gellert, M. and Lambowit, A., eds), pp. 813–835. American Society of Microbiology Press, Washington, DC.

Eickbush, T.H. and Eickbush, D.G. (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**: 477–485.

Ghesini, S., Luchetti, A., Marini, M. and Mantovani, M. (2011) The non-LTR retrotransposon R2 in termites (Insecta, Isoptera): characterization and dynamics. *J Mol Evol* **72**: 296–305.

Hickey, D.A. (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**: 519–531.

Hiruta, C., Nishida, C. and Tochinai, S. (2010) Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. *Chromosome Res* **18**: 833–840.

Kagramanova, A.S., Korolev, A.L. and Mukha, D.M.V. (2010) Analysis of the inheritance patterns of 5′-truncated copies of the german cockroach R2 retroposons in individual crosses. *Russian J Genet* **46**: 1290–1294.

Kazazian, H.H. (2004) Mobile elements: drivers of genome evolution. *Science* **303**: 1626–1632.

Kojima, K.K. and Fujiwara, H. (2005) Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22**: 2157–2165.

Kraaijeveld, K., Zwanenburg, B., Hubert, B., Vieira, C., de Pater, S., van Alphen, J.J.M. *et al.* (2012) Transposon proliferation in an asexual parasitoid. *Mol Ecol* **21**: 3898–3906.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lockton, S. and Gaut, B.S. (2010) The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol Biol* **10**: 10.

Loewe, L. and Lamatsch, D.K. (2008) Quantifying the threat of extinction from Muller's ratchet in the diploid Amazon molly (*Poecilia formosa*). *BMC Evol Biol* **19**: 88.

Luchetti, A. and Mantovani, B. (2013) Non-LTR R2 element evolutionary patterns: phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages. *PLoS ONE* **8**: e57076.

Luchetti, A., Cesari, M., Carrara, G., Cavicchi, S., Passamonti, M., Scali, V. *et al.* (2003) Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). *J Mol Evol* **56**: 587–596.

Malik, H.S. and Eickbush, T.H. (2000) NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. *Genetics* **154**: 193–203.

Malik, H.S., Burke, W.D. and Eickbush, T.H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16**: 793–805.

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**: 157–163.

Mingazzini, V. (2011) Transposable elements: structure and dynamic of the autonomous retrotransposon R2 in Arthropoda with non-canonical reproduction. PhD thesis, University of Bologna.

Mingazzini, V., Luchetti, A. and Mantovani, B. (2011) R2 dynamics in *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution. *Heredity* **106**: 567–575.

Nuzhdin, S.V. and Petrov, D.A. (2003) Transposable elements in clonal lineages: lethal hangover from sex. *Biol J Linn Soc* **79**: 33–41.

O'Donnell, K.A. and Burns, K.H. (2010) Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mobile DNA* **1**: 21.

Omilian, A.R., Cristescu, M.E.A., Dudycha, J.L. and Lynch, M. (2006) Ameiotic recombination in asexual lineages of *Daphnia*. *Proc Natl Acad Sci USA* **103**: 18638–18643.

Pérez-Gonzalez, C.E. and Eickbush, T.H. (2001) Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans*. *Genetics* **158**: 1557–1567.

Plohl, M., Luchetti, A., Meštrovic′, N. and Mantovani, B. (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* **409**: 72–82.

Ricci, A., Scali, V. and Passamonti, M. (2008) The IGS-ETS in *Bacillus* (Insecta Phasmida): molecular characterization and the relevance of sex in ribosomal DNA evolution. *BMC Evol Biol* **8**: 278.

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.

Schaack, S., Pritham, E.J., Wolf, A. and Lynch, M. (2010) DNA transposon dynamics in populations of *Daphnia pulex* with and without sex. *Proc R Soc Lond B* **277**: 2381–2387.

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.

Valizadeh, P. and Crease, T.J. (2008) The association between breeding system and transposable element dynamics in *Daphnia pulex*. *J Mol Evol* **66**: 643–654.

Wright, S. and Finnegan, D. (2001) Genome evolution: sex and transposable elements. *Curr Biol* **11**: R296–R299.

Zhang, X. and Eickbush, T.H. (2005) Characterization of active R2 retrotransposition in the rDNA locus of *Drosophila simulans*. *Genetics* **170**: 195–205.

Zhang, X., Zhou, J. and Eickbush, T.H. (2008) Rapid R2 retrotransposition leads to the loss of previously inserted copies via large deletions of the rDNA locus. *Mol Biol Evol* **25**: 229–237.

Zhou, J. and Eickbush, T.H. (2009) The pattern of R2 retrotransposon activity in natural populations of *Drosophila simulans* reflects the dynamic nature of the rDNA locus. *Plos Genet* **5**: e1000386.

Zhou, J., Eickbush, M.T. and Eickbush, T.H. (2013) A population genetic model for the maintenance of R2 retrotransposons in rRNA gene loci. *PLoS Genet* **9**: e1003179.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Biogeographic distribution of *Bacillus* stick insect taxa.

Figure S2. (A) Scenario inferred for the all.he simulation with unequal recombination in both parental individuals. Post-selection insertion frequencies are significantly different from those calculated for no recombination and post-recombination ($G = 14.5$; d.f. = 3; $P < 0.01$). (B) Scenario inferred for the one.ho simulation with unequal recombination in the heterozygous parental. (C) Scenario inferred for the one.ho simulation with unequal recombination in the homozygous parental. Post-recombination insertion frequencies are significantly different from those calculated for no recombination ($G = 29.8$; d.f. = 3; $P < 0.001$). Post-selection insertion frequencies are not significantly different from those calculated for post recombination ($G = 0.97$; d.f. = 3; $P = 0.241$). (D) Scenario inferred for the one.ho simulation with unequal recombination in both parental individuals. Post-recombination insertion frequencies are significantly different from those calculated for no recombination ($G = 20.4$; d.f. = 3; $P < 0.001$). Post-selection insertion frequencies are not significantly different from those calculated for post recombination ($G = 0.07$; d.f. = 3; $P = 0.105$).

Table S1. Goodness-of-fit G test of observed R2 insertion frequencies to Mendelian expectation under the three scenarios all.ho, all.he and one.ho for the bisexual offspring (see text).

Table S2. Primers used for R2 PCR amplification and sequencing.

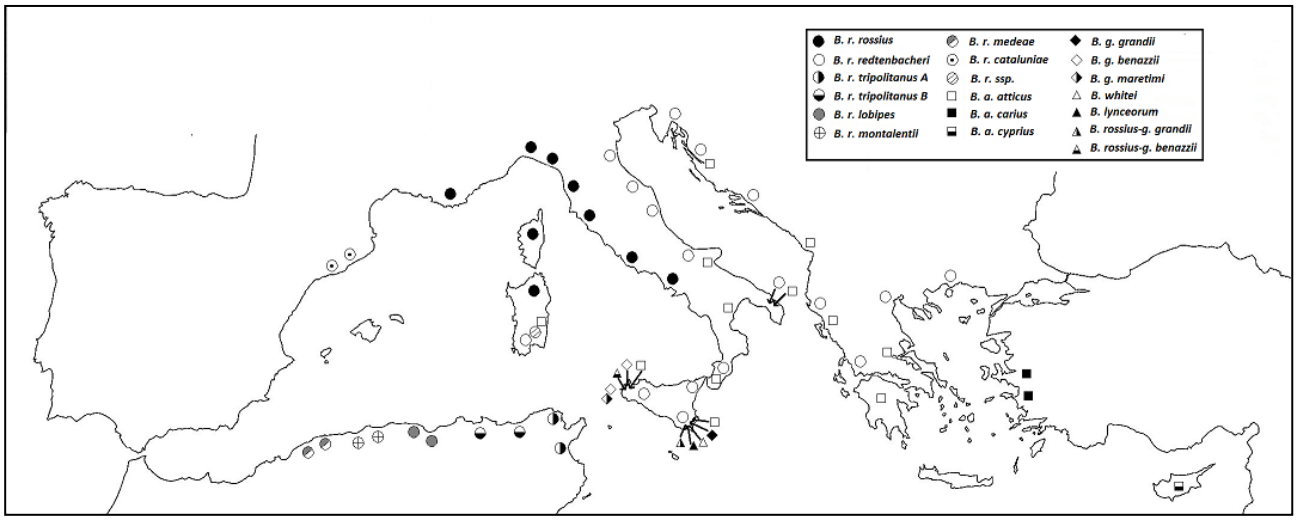


Figure S1. Biogeographic distribution of *Bacillus* stick insect taxa.

Figure S2A. Scenario inferred for the all.he simulation with unequal recombination in both parentals. Post-selection insertion frequencies are significantly different from those calculated for no recombination and post-recombination (G = 14.5; d.f. = 3; p < 0.01).

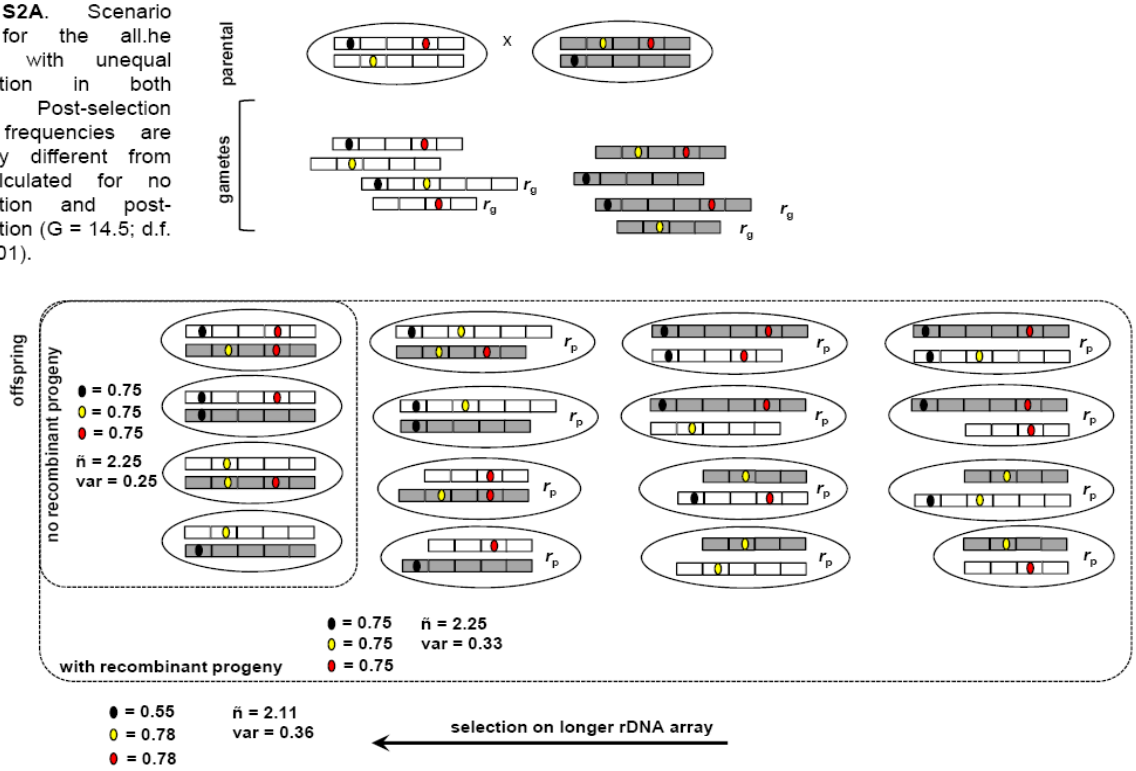


Figure S2B. Scenario inferred for the one.ho simulation with unequal recombination in the heterozygous parental.

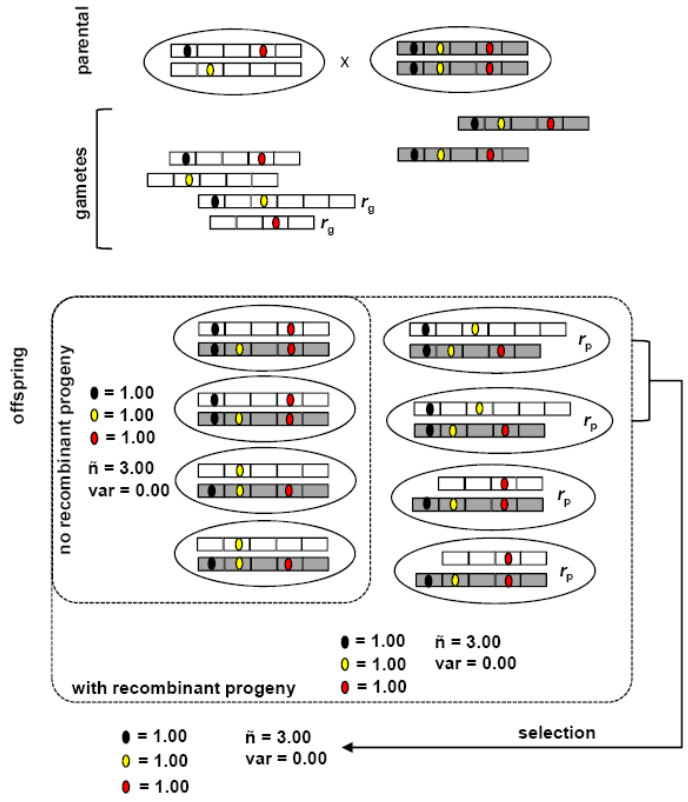


Figure S2C. Scenario inferred for the one.ho simulation with unequal recombination in the homozygous parental. Post-recombination insertion frequencies are significantly different from those calculated for no recombination (G = 29.8; d.f. = 3; p < 0.001). Post-selection insertion frequencies are not significantly different from those calculated for post recombination (G = 0.97; d.f. = 3; p = 0.241).

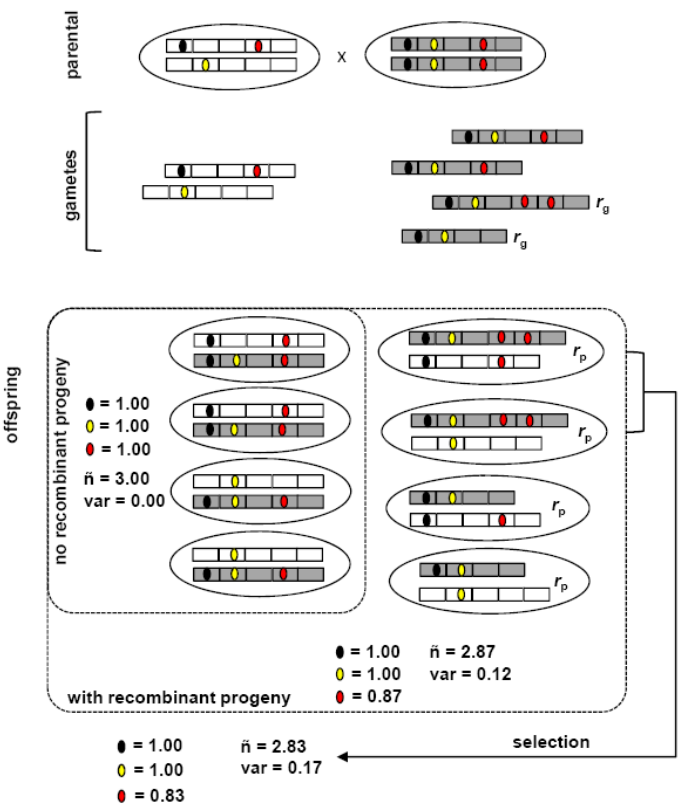


Figure S2D. Scenario inferred for the one.ho simulation with unequal recombination in both parentals. Post-recombination insertion frequencies are significantly different from those calculated for no recombination (G = 20.4; d.f. = 3; p < 0.001). Post-selection insertion frequencies are not significantly different from those calculated for post recombination (G = 0.07; d.f. = 3; p = 0.105).

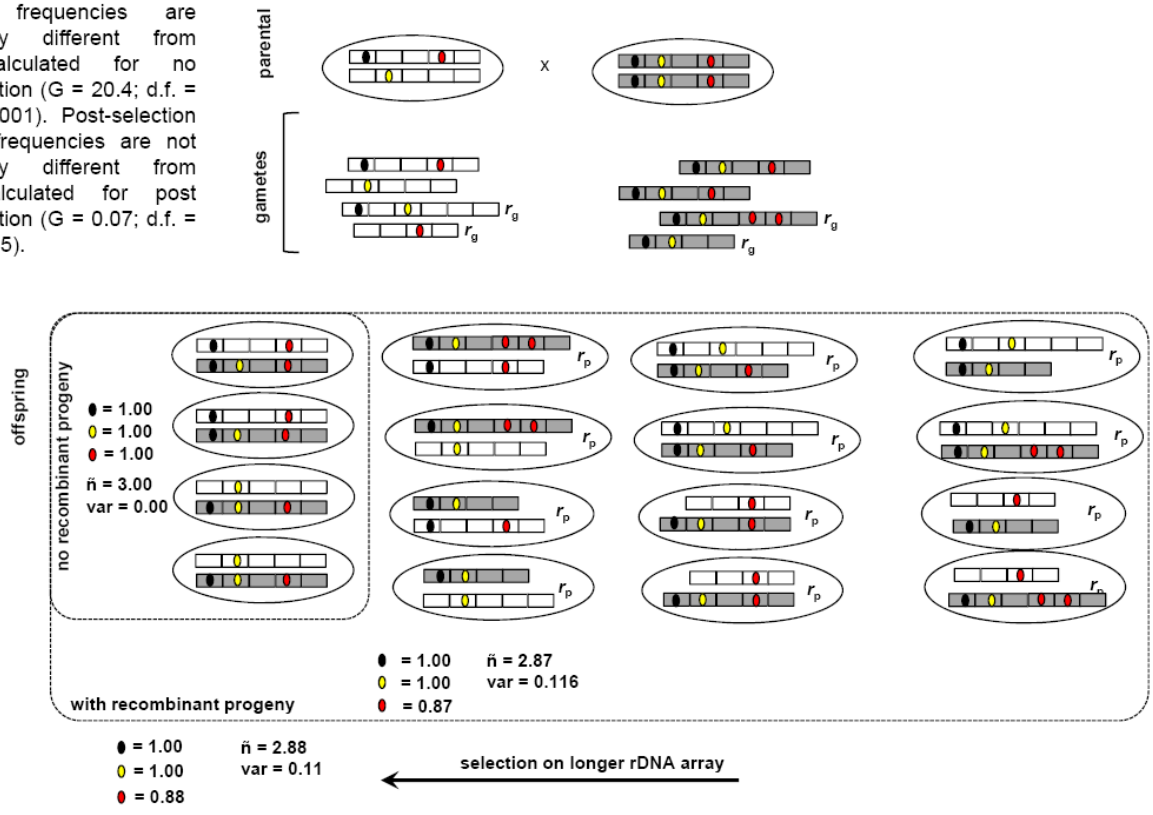


Table S1. Goodness-of-fit G test of observed R2 insertion frequencies to Mendelian expectation under the three scenarios all.ho, all.he and one.ho for the bisexual offspring (see text).

Crosses	Expected frequency sets	G (d.f.)	p
reCUR-♀5 × roANZ-♂15	all.ho	1640.9 (11)	< 0.0001
	all.he	385.4 (11)	< 0.0001
	one.ho	385.4 (11)	< 0.0001
reCUR-♀6 × roANZ-♂9	all.ho	214.7 (4)	< 0.0001
	all.he	214.7 (4)	< 0.0001
	one.ho	214.7 (4)	< 0.0001
roANZ-♀21 × roANZ-♂8	all.ho	4283.1 (19)	< 0.0001
	all.he	115.9 (19)	< 0.0001
	one.ho	1202 (19)	< 0.0001
roCAP-♀1 × roCAP-♂4	all.ho	1468.3 (15)	< 0.0001
	all.he	16.53 (15)	0.063
	one.ho	451.84 (15)	< 0.0001
roCAP-♀2 × roCAP-♂3	all.ho	1931.8 (12)	< 0.0001
	all.he	48.95 (12)	< 0.0001
	one.ho	307.00 (12)	< 0.0001

Table S2. Primers used for R2 PCR amplification and sequencing.

Primer name	Sequence 5' -> 3'	Reference
R2IF1	AAGCARGGNGAYCCNCTNTC	Kojima and Fujiwara (2005)
28S-F2	GAATCCGACTGTCTAATTAAAACAAAG	Mingazzini et al. (2011)
28S-R	TCCATTGCTGCGCGTCACTAATTAGATGAC	this study
rR2rin	GACTGTCCAACAATAGGAGGGAAT	this study
rR2rin2	CACCAGGAGATTAGTTTGGTTTCT	this study
rR2din2	GCATGTCCAAGGATAAAGTCTAAAA	this study
rR2din	AACGACTATCAGTTCGTTGAATAGG	this study
rR2Prin	GACTGTCCAACAATAGGAGGGAAT	this study
rR2Pdin	GCATGTCCAAGGATAAAGTCTAAAA	this study
rR2Prin2	CACCAGGAGATTAGTTTGGTTTCT	this study

2.1.2. Degenerate R2 Element Replication and rDNA Genomic Turnover in the *Bacillus rossius* Stick Insect

Having found a degenerate R2 variant in the *B. rossius* genome we wanted investigate the mechanism underlying R2 degenerate variant maintenance in the genome. The R2 5' ends have been sequenced from two specimens of *B. rossius rossius* and one specimen of *B. rossius redtenbacheri* through PCR amplification and cloning. The analyses were performed on 142 clones and yielded 74 full-length elements (i. e. without 5' end deletions) and 68 truncated variants (5' end deletions ranging from 101 bp to 1297 bp). These analyses showed the presence of the degenerate variant (R2Br^{deg}) even in the *B. rossius rossius* samples. The 28S/5' R2 junction sequences and the ability of the ribozyme to self-cleave the 28S/R2 co-transcript and, thus, to produce a mature R2Br RNA were analyzed. These analyses, performed with the collaboration of the Professor Danna G. Eickbush (University of Rochester, New York, USA), demonstrated a "G" to "A" substitution in the 28S gene that occurs in all sequenced degenerate element junctions and the presence of additional 23 bp in the R2Br^{deg} 5' UTR that appears to interfere with the correct formation of the ribozyme. It follows that the ability of the ribozyme to self-cleave is very low. However, the presence of R2^{deg} in *B. rossius tripolitanus* A (Bonandin et al., 2014) indicates that this degenerate element was already present in the ancestral *B. rossius* genome, when the separation of Europe and North Africa led to allopatric *B. rossius* lineage break (>5 Myr ago; Mantovani et al., 2001). Therefore, there should be a mechanism that allows the maintenance of these elements even if unable to encode a protein or to self-cleave. The most plausible explanation is given by genomic turnover mechanisms, and therefore concerted evolution involving 28S ribosomal genes.

This part of my work is presented in the paper:

Martoni F., Eickbush D. G., Scavariello C., Luchetti A. and Mantovani B. 2015. Dead element replicating: degenerate R2 element replication and rDNA genomic turnover in the *Bacillus rossius* stick insect (Insecta: Phasmida)**.** PloS One, Mar 23;10(3):e0121831.

PLOS | ONE

RESEARCH ARTICLE

Dead Element Replicating: Degenerate R2 Element Replication and rDNA Genomic Turnover in the *Bacillus rossius* Stick Insect (Insecta: Phasmida)

Francesco Martoni[1¤], Danna G. Eickbush[2], Claudia Scavariello[1], Andrea Luchetti[1]*, Barbara Mantovani[1]

1 Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Bologna, Bologna, Italy,
2 Department of Biology, University of Rochester, Rochester, New York, United States of America

¤ Current Address: Bio-Protection Research Centre, Faculty of Agriculture & Life Sciences, Lincoln University, Christchurch, New Zealand
* andrea.luchetti@unibo.it

Citation: Martoni F, Eickbush DG, Scavariello C, Luchetti A, Mantovani B (2015) Dead Element Replicating: Degenerate R2 Element Replication and rDNA Genomic Turnover in the *Bacillus rossius* Stick Insect (Insecta: Phasmida). PLoS ONE 10(3): e0121831. doi:10.1371/journal.pone.0121831

Academic Editor: Jürgen Schmitz, University of Muenster, GERMANY

Received: November 6, 2014

Accepted: February 4, 2015

Published: March 23, 2015

Data Availability Statement: Sequence data are available from the GenBank database (accession numbers KP657751-KP657892).

Funding: This work was supported by RFO - University of Bologna and Canziani funding to AL and BM and National Institutes of Health Grant Number R01GM42790 to DGE. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

R2 is an extensively investigated non-LTR retrotransposon that specifically inserts into the 28S rRNA gene sequences of a wide range of metazoans, disrupting its functionality. During R2 integration, first strand synthesis can be incomplete so that 5' end deleted copies are occasionally inserted. While active R2 copies repopulate the locus by retrotransposing, the non-functional truncated elements should frequently be eliminated by molecular drive processes leading to the concerted evolution of the rDNA array(s). Although, multiple R2 lineages have been discovered in the genome of many animals, the rDNA of the stick insect *Bacillus rossius* exhibits a peculiar situation: it harbors both a canonical, functional R2 element (R2Br^{fun}) as well as a full-length but degenerate element (R2Br^{deg}). An intensive sequencing survey in the present study reveals that all truncated variants in stick insects are present in multiple copies suggesting they were duplicated by unequal recombination. Sequencing results also demonstrate that all R2Br^{deg} copies are full-length, i. e. they have no associated 5' end deletions, and functional assays indicate they have lost the active ribozyme necessary for R2 RNA maturation. Although it cannot be completely ruled out, it seems unlikely that the degenerate elements replicate via reverse transcription, exploiting the R2Br^{fun} element enzymatic machinery, but rather via genomic amplification of inserted 28S by unequal recombination. That inactive copies (both R2Br^{deg} or 5'-truncated elements) are not eliminated in a short term in stick insects contrasts with findings for the *Drosophila* R2, suggesting a widely different management of rDNA loci and a lower efficiency of the molecular drive while achieving the concerted evolution.

Introduction

Transposable elements (TEs) are DNA sequence units able to move within the genome. TEs constitute a significant fraction, or even the majority, of some eukaryotic genomes, the percentage reaching 77% in *Rana esculenta* and 85% in *Zea mays* [1]. Their diversity is increasing with new families being continuously discovered, especially as the large number of sequenced genomes is analyzed. The role of transposable elements in evolution is highly debated, but their effects range from beneficial to negative, obviously owing to their impact on host fitness [1, 2, 3].

Class I TEs comprises mobile elements whose movement requires the activity of a reverse transcriptase. A major subclass is represented by non-long terminal repeat (non-LTR) retrotransposons [1]; among them, R2 is one of the most investigated elements and serves as a model for understanding the non-LTR retrotransposition mechanisms. Its structure comprises a single open reading frame (ORF) flanked by two untranslated regions (UTR); the ORF encompasses a central reverse transcriptase (RT) domain which includes RNA binding motifs [4], the DNA-binding motifs at the N-terminus and the endonuclease domain (EN) at the C-terminus. The protein C-terminal end has a cysteine-histidine (zinc finger) motif (CCHC) while the N-terminal domain can contain one (CCHH), two (CCHH + CCHH or CCHC + CCHH), or three (CCHH + CCHC + CCHH) zinc finger motifs ([5] and references therein). R2 has strict sequence specificity for an insertion target site in the 28S rRNA gene (rendering the gene non functional) and it occurs in a wide range of animal taxa, from diploblastic organisms to lower vertebrates [5, 6]. Evolutionarily speaking, R2 belongs to an ancient group of retrotransposons whose members insert specifically into tandem repeats, although a few exceptions have been found [7]. This might represent an adaptive strategy to escape genome purging by limiting damage to a subset of the functional genes among the redundant copies [8, 9].

The R2 mechanism of integration requires a 3' hydroxyl group at a DNA break to prime reverse transcription (target primed reverse transcription, TPRT [10]). Although the reverse transcriptase occasionally fails to reach the 5' end of the RNA template, a complete integration event can still take place but the result is a 5' end truncated copy. The location of the truncation is typically unique; therefore, this length variation at the R2 5' end can be used to evidence and track the element activity [11, 12].

The R2 RNA template is produced by co-transcription with the rDNA unit followed by self-cleavage. The 5' end of the R2 elements, in fact, can fold into structures very similar to the self-cleaving ribozymes encoded by the Hepatitis Delta Virus (HDV) [13, 14]. These structures are capable of self-cleavage as demonstrated for the R2 elements of many species with cleavage of the co-transcript occurring upstream of the 28S/R2 5' junction in many species (for example, in the earwig *Forficula auricularia*) or at the junction of the 28S gene and the 5' end of the element in some species (for example, in the fruit fly *Drosophila melanogaster*) [15]. This dichotomy in the location of self-cleavage has been correlated with the types of R2 junctions within a species. R2 5' junctions are uniform for most R2s in which self-cleavage is upstream in the rRNA sequences but they are variable for most R2s in which cleavage is at the R2 5' end. It has been postulated that the presence of 28S sequences allows the annealing of the first DNA strand synthesized during retrotransposition to the target site and uniformly primes second strand synthesis; in the absence of 28S sequences, priming depends on chance microhomologies between the target site and the first DNA strand [16, 17].

Owing to its location in the array, R2 dynamics is affected by molecular drive which shapes the composition of the rDNA locus. Molecular drive includes a variety of genomic turnover mechanisms (unequal crossing-over, gene conversion, rolling circle replication, etc.) that determines the spread of new units within the same genome (homogenization) and subsequently in

the population, through bisexual reproduction (fixation) [18, 19]. This variability pattern is also known as concerted evolution [18].

We recently analyzed R2 in the stick insect species *Bacillus rossius* (R2Br). In addition to a canonical element encoding a 1054 amino acid sequence comprising all known R2 domains and a single ZF motif (CCHH type) at the N-terminal end (named R2Br^{fun}), a degenerate but closely related (9.2% nucleotide divergence) element has been also isolated, R2Br^{deg} [12]. This latter element exhibits 14 frameshift mutations and one stop codon within the open reading frame, and it is at least 5 Myr old as the degenerate element is found in the Italian subspecies *B. rossius rossius* and *B. rossius redtenbacheri* and also in the North-African *B. rossius tripolitanus* A [12, 20].

A population sequencing survey, based on the 3' end of the R2Br element, indicated that all *B. r. rossius* samples host only R2Br^{fun}, while *B. r. redtenbacheri* populations had either only one element variant (R2Br^{deg} or R2Br^{fun}) or both variants in different proportions. Interestingly, no relationships emerged between the presence/absence of a particular R2Br variant and the reproductive strategies (bisexual *vs* parthenogenetic). On the other hand, tracking element activity in these subspecies revealed new R2Br insertions even in the populations showing only R2Br^{deg} in the sequencing survey [12]. Moreover, sequence data clearly indicated a mutation pattern of R2Br^{deg} consistent with an ongoing replicative activity. We, therefore, suggested that R2Br^{deg} could either represent a non-autonomous element that exploits the retrotransposition machinery of an R2Br^{fun} not identified in our survey or duplicates along with the host 28S sequences through genomic turn over mechanisms.

In this paper, we delve further into this issue to better understand the mechanisms underlying the R2Br^{deg} duplication. We, therefore, analyze in the genomes of three *B. rossius* populations three features that have been linked to the retrotransposition activity of functional R2 elements: i) the sequence of the 28S/5' R2 junction of both full-length and truncated elements, ii) the potential to fold the 5' junction sequences into a HDV-like ribozyme structure, and iii) the ability of the ribozyme to self-cleave the 28S/R2 co-transcript and, thus, to produce a typical mature R2Br RNA.

Materials and Methods

Sampling and DNA isolation

Specimens have been collected in areas where specific permission for sampling is not requested, as sampling sites are located in public areas with no restrictive or protection laws enforced. Animals sampled are not endangered or protected species. Individuals of *B. r. rossius* from Capalbio (GR, Tuscany; roCAP; one male and one female) and Anzio (RM, Lazio; roANZ; two females), and of *B. r. redtenbacheri* from Patti (ME, Sicily; rePAT; one male and two females) were field collected and frozen at -80°C until molecular analysis. To assure no kinship between the analyzed insects, when possible, specimens were chosen either from different sampling years (Anzio) or from collection sites located at the opposite sides of the same sampling area (Capalbio). Total DNA was extracted from a single stick insect leg or from the whole body with the standard phenol/chloroform protocol. A previous R2Br survey [12] was carried out on the same sampling; moreover, the presently analyzed roCAP female and one female of rePAT are the same specimens used.

R2 elements isolation, sequencing and analysis

R2 5' ends were PCR amplified using a primer anchored to the 28S rRNA gene, 64 bp upstream of the R2 insertion site (28SF2: `5'-GTCAAAGTGAAGAAATTCAACGAAG-3'`), coupled with

two primers anchored inside R2: starting either at base 1917 (RR2Rin: 5'-CCATTCCATT-CAATACAGTATCTCC-3') or at base 1424 (R21424r: 5'-AAGCCCAAACAGCAGACGGC-3').

PCR products were ~2000 bp or ~1400 bp long (with 28SF2+RR2Rin or 28SF2+R21424r, respectively) when the full-length element was amplified (i.e., no 5' end deletions occurred); also shorter amplicons were produced, and these represented truncated variants whose length depended on the extent of the 5' end deletion.

PCR amplifications were performed in a 50-μl reaction mix using the GoTaq G2 Flexi kit (Promega), following the manufacturer's instructions. Thermal cycling was as follows: initial denaturation step at 94°C for 5 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 48°C for 30 s and extension at 72°C for 2 min, and a final extension at 72°C for 7 min.

Amplicons were run on a 1.5% agarose gel and bands were eluted from the gel using the Wizard SV Gel and PCR Clean-Up System kit (Promega). Fragments were, then, inserted into a pGEM-T Easy vector (Promega) and used to transform *E. coli* DH5α cells. Recombinant colonies were PCR-amplified with T7/SP6 primers and sequenced at Macrogen Inc.—Europe Lab. Sequence data are available in GenBank under the acc. nos. KP657751-KP657892.

Sequence alignment with Clustal W algorithm and pairwise genetic divergence (p-distance) were calculated with MEGA v. 6 [21]; Tajima's Ds have been calculated using DnaSP v. 5.1 [22]. The phylogenetic inference has been carried out using MrBayes 3.2.2 [23], setting the GTR model of substitution. The Markov Chain Monte Carlo process was set on two simultaneous tree searches running for 10^6 generations and tree sampling every 500 generations. Runs' convergence was assessed through the variance of split frequencies (< 0.01), PSRF ≥ 1.00 and ESS ≥ 200, after a conservative *burn-in* period of 25%.

DNA templates for T7 co-transcription/cleavage reactions

DNA templates for RNA transcription were generated by PCR amplification of cloned R2 junctions from a specimen of *B. r. rossius* from Anzio (both functional and degenerate copies) with unincorporated primers and nucleotides removed using a PCR Purification Kit (BioBasics). The specific primers used can be found in the S1 Table. Self-cleavage was assayed as previously described [13]. In short, PCR templates were incubated in transcription buffer with 20 units of T7 RNA Polymerase (Invitrogen) and trace amounts of [α-^{32}P]UTP for one hour at 42°C, the reactions stopped on ice by the addition of 4 volumes of 95% formamide, and the denatured RNA products separated on 8M urea, 5% acrylamide gels. After fixing and drying, the gels were exposed to a phosphorimager screen and analyzed using QuantityOne (BioRad).

Results

R2 sequences analysis

As a first step to determine how the elements were duplicating in stick insects, the R2 5' ends were acquired from two specimens of *B. r. rossius* and one specimen of *B. r. redtenbacheri* by PCR amplification and cloning. The analysis of 142 clones yielded 74 full-length elements (i. e. without 5' end deletions) and 68 truncated variants (5' end deletions ranging from 101 bp to 1297 bp) (Table 1). Over 90% of these clones represents different copies as they had unique sequences, while 13 clones had sequences identical to other clones (Fig. 1).

Truncated variants were largely sample-specific, the only exception being the 1297 bp truncation shared between roCAP and rePAT populations (Table 1). The dataset was then analyzed with a phylogenetic method, adding also the reference sequences R2Br^{fun} (GenBank acc. no. KJ958674) and R2Br^{deg} (acc. no. KJ958675 [12]). The resultant cladogram shows mainly polytomic terminal branches but well-structured clustering at the deepest nodes (Fig. 2). Two main clusters emerge from this analysis: one includes the R2Br^{deg} reference element and

Table 1. Distribution of sequenced R2Br full-length and truncated variants.

5' end deletion (bp)	roANZ		roCAP		rePAT	
	fun	*deg*	*fun*	*deg*	*fun*	*deg*
full-length	4	9	6	17	4	34
-101	10	0				
-714			8	0		
-1062	20	0				
-1172			10	0		
-1297			10	0	10[a]	0

Sequenced variants (indicated by the extent of their 5' end deletion with respect to the consensus), either functional (*fun*) or degenerated (*deg*), distribution in the three analyzed populations (*B. r. rossius*; Anzio: roANZ; Capalbio: roCAP. *B. r. redtenbacheri*; Patti: rePAT).

[a] Including recombinant elements.

doi:10.1371/journal.pone.0121831.t001

34 full-length elements from rePAT, 17 from roCAP and nine from roANZ; the other cluster embodies the R2Br^{fun} reference sequence, the remaining full-length elements and most of the truncated variants. The discovery of the degenerate R2 variant in roCAP and roANZ as well as of the functional R2 variant in rePAT was unexpected since in our previous analysis the sequencing of the element's 3' end did not indicate the presence of these variants in these same populations [12]. It is likely that the different primer pairs used in this study perform better in

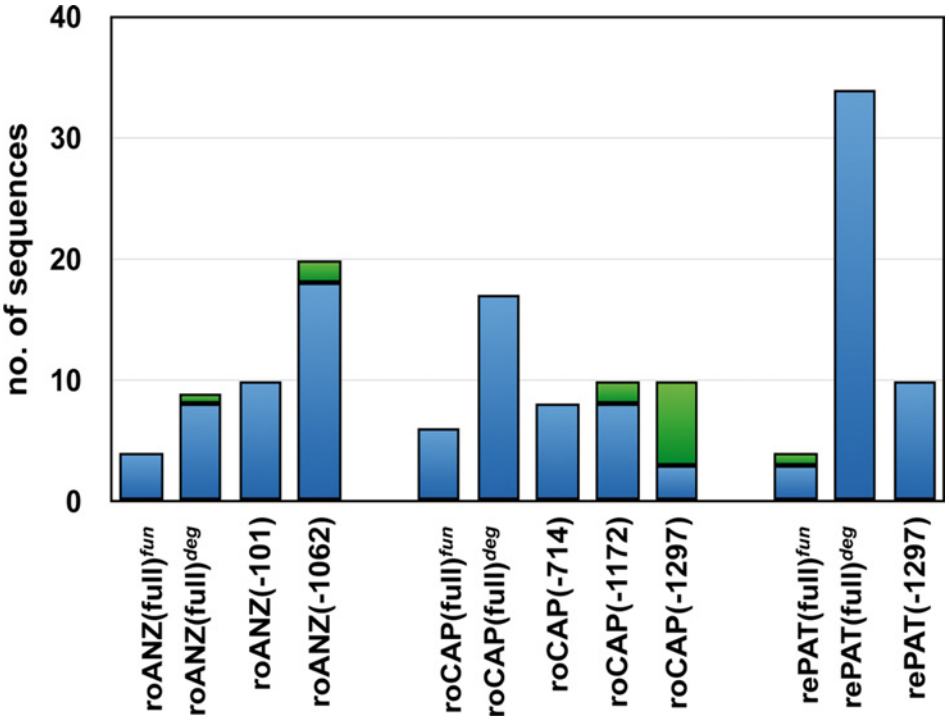


Fig 1. Proportion of unique sequence clones out of the number of sequenced clones. Bins indicate the number of sequenced R2 variants (full-length and truncations) per sample, as listed in Table 1. The blue part of each bin indicates the proportion of unique sequences, i.e. those differing from the others by at least one nucleotide substitution. The green part of each bin represents the proportion of sequences identical to another one.

doi:10.1371/journal.pone.0121831.g001

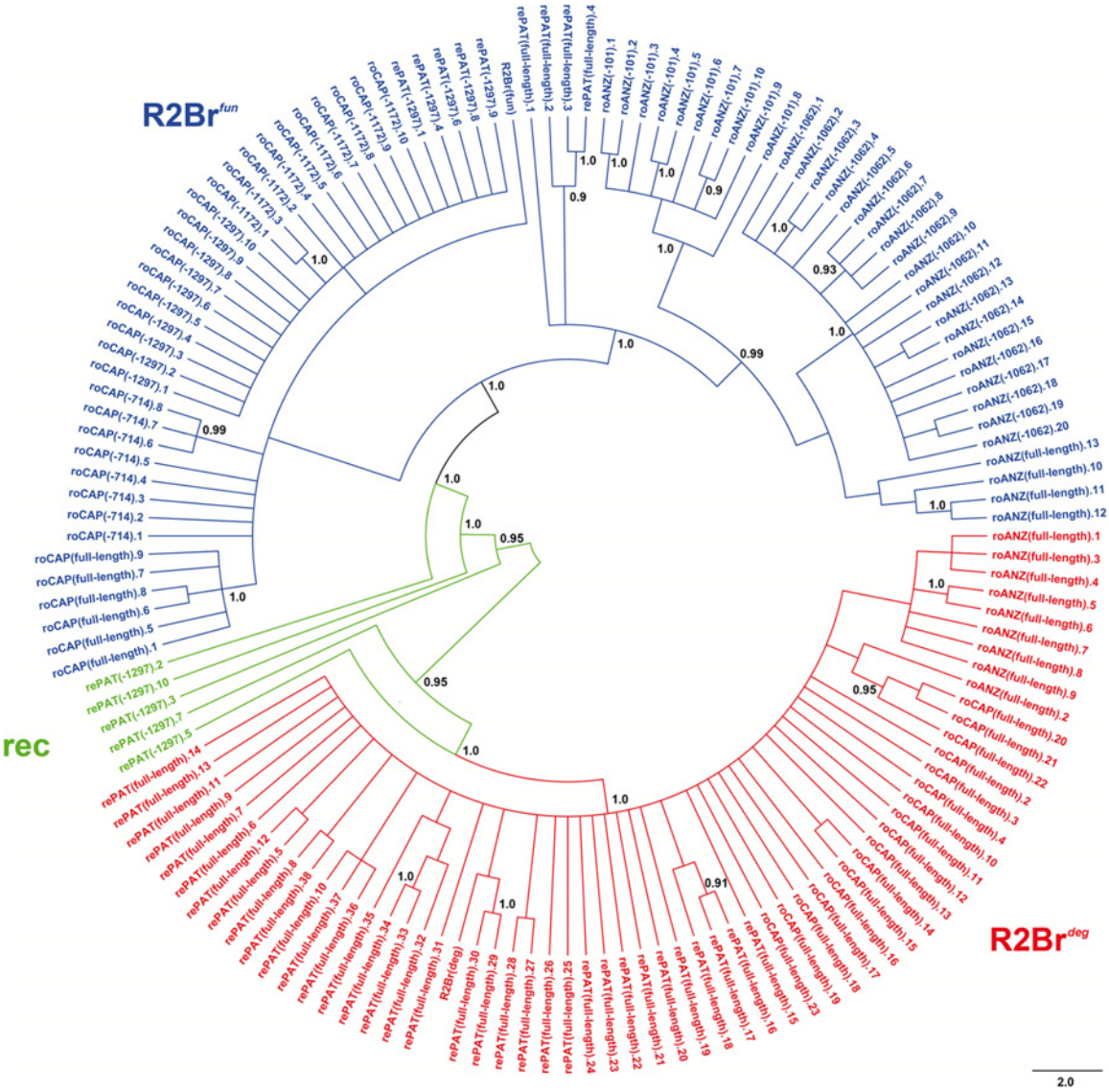


Fig 2. Bayesian phylogeny of the full-length and truncated elements (-ln L = 12418.90). Numbers at nodes indicate posterior probabilities; only values ≥ 0.90 are reported. R2Br^{fun} and R2Br^{deg} reference sequences are also included as R2Br(fun) and R2Br(deg). The sequences indicated with "rec" are those identified as recombinant between R2Br^{fun} and R2Br^{deg} (see Fig. 3).

doi:10.1371/journal.pone.0121831.g002

sampling R2 within the analyzed genomes; moreover, it is to be noted that present samplings involve a 4–5-fold higher number of sequences per analyzed population. While the "degenerate" cluster does not show any sub-structuring, the "functional" cluster exhibits a clear separation between roCAP and roANZ samples with most of the rePAT sequences being intermingled among the other two groups (Fig. 2).

Five rePAT truncated sequences, however, do not clearly fall within either of the two major clusters (Fig. 2). Sequence inspection of diagnostic nucleotides characterizing R2Br^{fun} and R2Br^{deg} reference sequences revealed that these five sequences are recombinants between the two R2 variants (Fig. 3). More precisely, the five rePAT truncated elements showed R2Br^{fun} diagnostic nucleotides at their 5' and 3' ends and, to a different extent, R2Br^{deg} diagnostic bases in the internal region. It is likely that these recombinants are the result of gene conversion rather

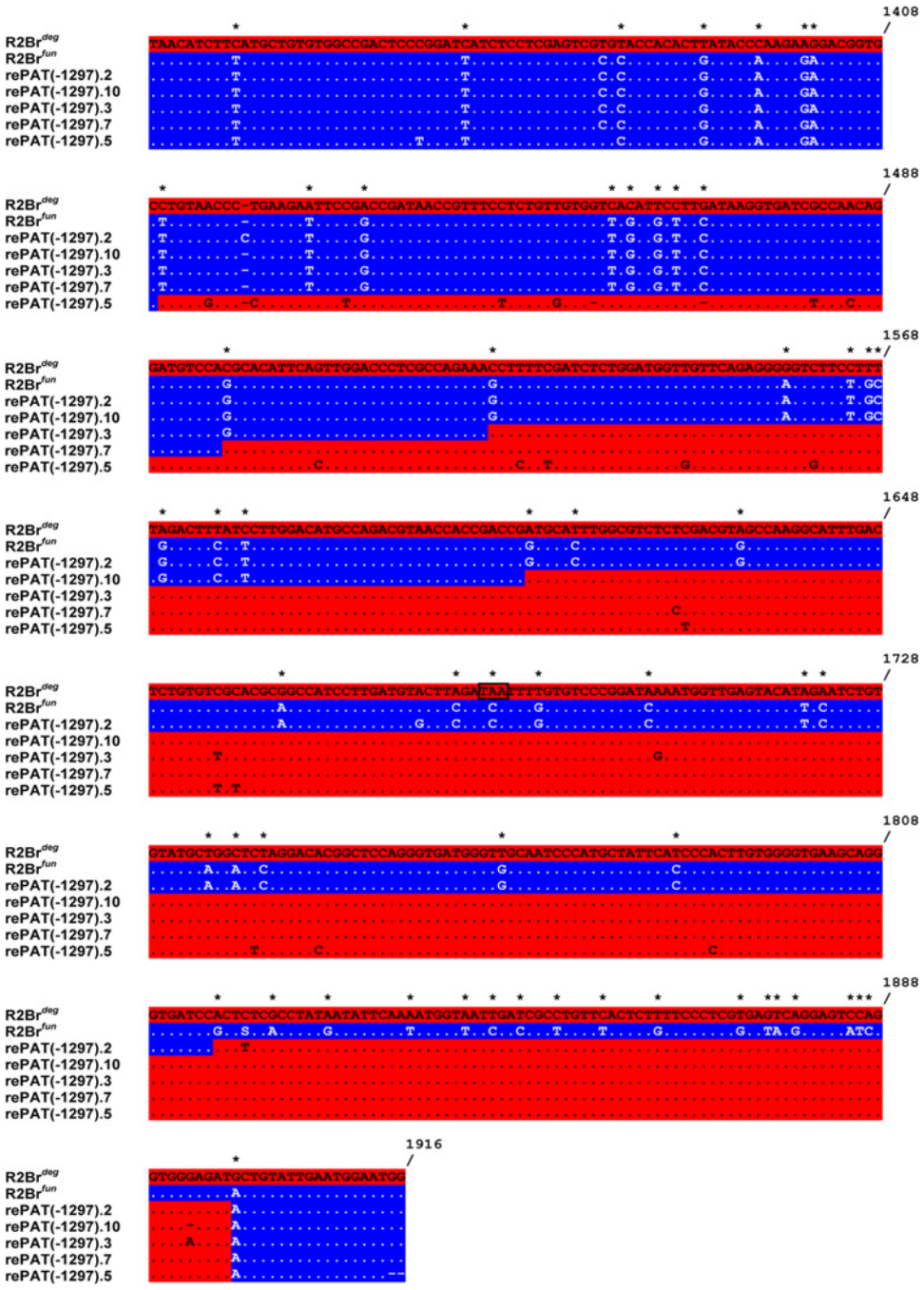


Fig 3. Alignment of recombinant R2Br sequences with the reference sequences, R2Br^{deg} and R2Br^{fun}. Diagnostic sites for distinguishing the two R2Br variants are indicated by asterisks. Different colors shading indicates whether the nucleotide sequence belongs to the R2Br^{deg} (red) or R2Br^{fun} (blue) element. Numbers at the end of lines refer to nucleotide positions relative to the alignment including the two full-length consensus degenerate and functional sequences. The R2Br^{deg} internal stop codon (TAA; [12]) is boxed.

doi:10.1371/journal.pone.0121831.g003

than template switching as the latter would require two jumps from the 5' end of one transcript to the middle of another transcript; such jumps were not observed during *in vitro* experiments with the R2 protein [24]. Also, the presence of this same truncation in roCAP suggests these recombinant copies were originally derived from the functional variant (Table 1; Fig. 2).

Table 2. Nucleotide variability and Tajima's D of sequenced R2Br.

	N	Overall	5' UTR	ORF	Tajima's D[a]
R2Br^{fun}	82	0.0199	0.0195	0.0198	-1,8064*
Full-length	14	0.0177	0.0263	0.0167	-0,4489^{ns}
Truncated	68	0.0208	n.a.[b]	0.0208	-2,2862**
R2Br^{deg}	60	0.0075	0.005	0.0078	-2,8239***
fun vs *deg*	142	0.109	0.153	0.108	

[a] Probability levels for Tajima's D statistical significance:

* $p < 0.05$;

** $p < 0.01$;

*** $p < 0.001$;

^{ns} not significant.

[b] The part of 5' UTR in the truncated elements dataset is covered only by 10 sequences from the same sample (roANZ) and represents only ~38% of the region: it has been, therefore, not considered.

doi:10.1371/journal.pone.0121831.t002

R2Br^{fun} sequence variability is quite high with respect to the R2Br^{deg} and the 5' end deleted variants are appreciably more variable than the full-length ones (Table 2). It is to be noted that several unique sequences were detected for all full-length (functional and degenerate) and 5'-truncated element types, the percentage of different sequences ranging from 30% to 100% (Fig. 1). The divergence between the functional and the degenerate R2Br is in line with previous estimates, as well as Tajima's D values [12]. Quite interestingly, the Tajima's D calculated only on the full-length R2Br^{fun} is not significantly different from zero (Table 2).

The 28S/R2 5' junctions and the autocatalytic ribozyme

As previously demonstrated for R2 elements in many species, the 5' end of the R2 RNA is processed from a 28S/R2 co-transcript via an encoded ribozyme [13, 14]. An analysis of the 5' junctions for the sequenced clones revealed several clues to the putative ribozyme structure(s) for R2Br (Table 3). First, full-length junctions for roANZ, roCAP, and rePAT are uniform suggesting the R2 ribozyme would cleave in the upstream 28S sequences of the co-transcript [15]. Second, while the junctions are also uniform for the degenerate copies, the presence of the non-consensus "A" in the upstream 28S sequences in each junction suggested that a putative ribozyme would have to cleave upstream of this nucleotide in order to regenerate the "A" at the DNA target site. Third, the nucleotide changes in all roCAP full-length junctions also suggested that if it encoded an active ribozyme, it would cleave upstream of these 28S gene changes.

A double pseudoknot structure, much like the secondary structures obtained for other insect R2s, could be generated using the sequences from the 5' junction of the full length R2Br^{fun} elements for roANZ and rePAT (Fig. 4A). A very similar secondary structure, albeit with a J1/2 loop which was 23 bp longer, seemed possible for the degenerate variant. Many of the nucleotide differences found in the degenerate element maintained base pairing in the P1, P2, and P4 stems which also suggested that at least at one time it encoded an active ribozyme (Fig. 4A). Consistent with the location of self-cleavage in many species, these structures suggest that an encoded ribozyme would self-cleave at a position 28 nucleotides upstream of the R2 insertion site.

These ribozyme structure predictions suggest that the roCAP full-length junctions (Table 3) are aberrant and would not encode a functional ribozyme. The junction likely originated from an insertion event in which a small portion of the 28S gene sequence at the 5' end of the RNA was "lost" during cDNA synthesis followed by the addition of non-templated nucleotides at the

Table 3. Sequenced 28S rRNA fragments upstream of the R2 insertion site.

28S sequence	Sample distribution	R2Br variant
GAAGCGCGGGTAAAC*GGCGGGA*GTAACTATGACTCTCTTAAGG	28S consensus	—
GAAGCGCGGGTAAAC*GGCGGGA*GTAACTATGACTCTCTTAA–↓	roANZ, rePAT (full-length); roANZ (-1062)	fun
GAAGCGCGGGTAA-C*ta-tGGA*GTAACTATGACTCTCTTAA–↓	roCAP (full-length)	fun
GAAGCGCGGGTAAAC*GGCGGGA*GTAACTATGACTCTCTT––↓	roANZ (-101); roCAP(-714)	fun
GAAGCGCGGGTAAAC*GGCGGGA*GTAACTATGACTCTC–––↓	roCAP (-1297)	fun
GAAGCGCGGGTAAAC*GGCGGGA*GTAACTATGACTCT–––-↓	rePAT (-1297)	fun
GAAGC––––––––––––––––––↓	roCAP (-1172)	fun
GAAGCGCGGGTAAAC*aGCGGGA*GTAACTATGACTCTCT––-↓	roANZ, roCAP, rePAT (full-length)	deg

The 28S junction sequences detected among the stick insects sampled with the associated R2Br variant (*fun*: functional; *deg*: degenerate) indicated at the far left. Nucleotide substitutions relative to the consensus sequence are in lower-case letters; deletions are denoted with dashes. Arrows mark the positional start of R2 sequences with the species and type (full-length or truncation length) of the element distribution indicated. The portion of the 28S sequence involved in the formation of the ribozyme P1 stem is in italic and underlined.

doi:10.1371/journal.pone.0121831.t003

start of second strand synthesis. The number of full-length elements sequenced is limited; therefore more sequencing and direct experiments are necessary to determine whether this stick insect population contains an active R2. Likewise, the junctions associated with the truncated elements are typical, i. e. they have variable deletions of upstream 28S sequences (2–38 nucleotides, Table 3), and each has at least part of—and roCAP(-1172) the entire—ribozyme secondary structure deleted so none would be able to self-cleave from a potential co-transcript.

R2Br^{fun} and R2Br^{deg} elements self-cleavage assay

To test for R2Br self-cleavage, templates comprising sequences from two 28S/R2 5' junctions for the functional element (which differed from each other by two nucleotides in the upstream 28S region and one in the L4 loop) and two degenerate element junctions (which differed from each other by a single substitution in the L4 loop) were generated by PCR amplification of cloned junctions. The T7 generated RNAs tested are diagrammed in Fig. 4B. An earwig R2 RNA was used as a positive control for self-cleavage and as a marker for the location of cleavage. As predicted, the RNAs from the R2Br^{fun} element showed self-cleavage which is consistent with a position 28 bp upstream of the R2 insertion site and at a level (average of 77%) comparable to that observed for the earwig ribozyme (86%) (Fig. 4C, lanes a-e). The RNAs derived from the degenerate element, however, had no detectable cleavage (Fig. 4C, lanes f-i).

To corroborate that the degenerate elements were incapable of self-cleavage from a 28S co-transcript, especially given the remarkably similar HDV-like structures envisioned using the functional and degenerate element sequences (Fig. 4A), mutant templates were generated. These new constructs addressed the effect on self-cleavage of specific sequence differences noted between the functional and degenerate variants. First a "G" to "A" substitution was introduced at the base of the P1 stem in the R2Br^{fun} element while at the homologous location an "A" to "G" substitution was introduced in the R2Br^{deg} element. Self-cleavage decreased almost 10 fold (76% to 8%) for the modified ribozyme from the functional element (Fig. 4C, lanes a and j) while the modified ribozyme associated with the degenerate element now showed a detectable level of self-cleavage (5%, Fig. 4C, lane k). The removal of a 41-nucleotide region which encompasses the 23 bp insertion, as well as multiple sequence differences with the functional J1/2 loop, from the 5' end of the degenerate element ribozyme also resulted in a comparable increase in self-cleavage (6%, Fig. 4C, lane l). Introducing both the "A" to "G" substitution in the P1 stem and the deletion in the J1/2 loop to the R2Br^{deg} element ribozyme

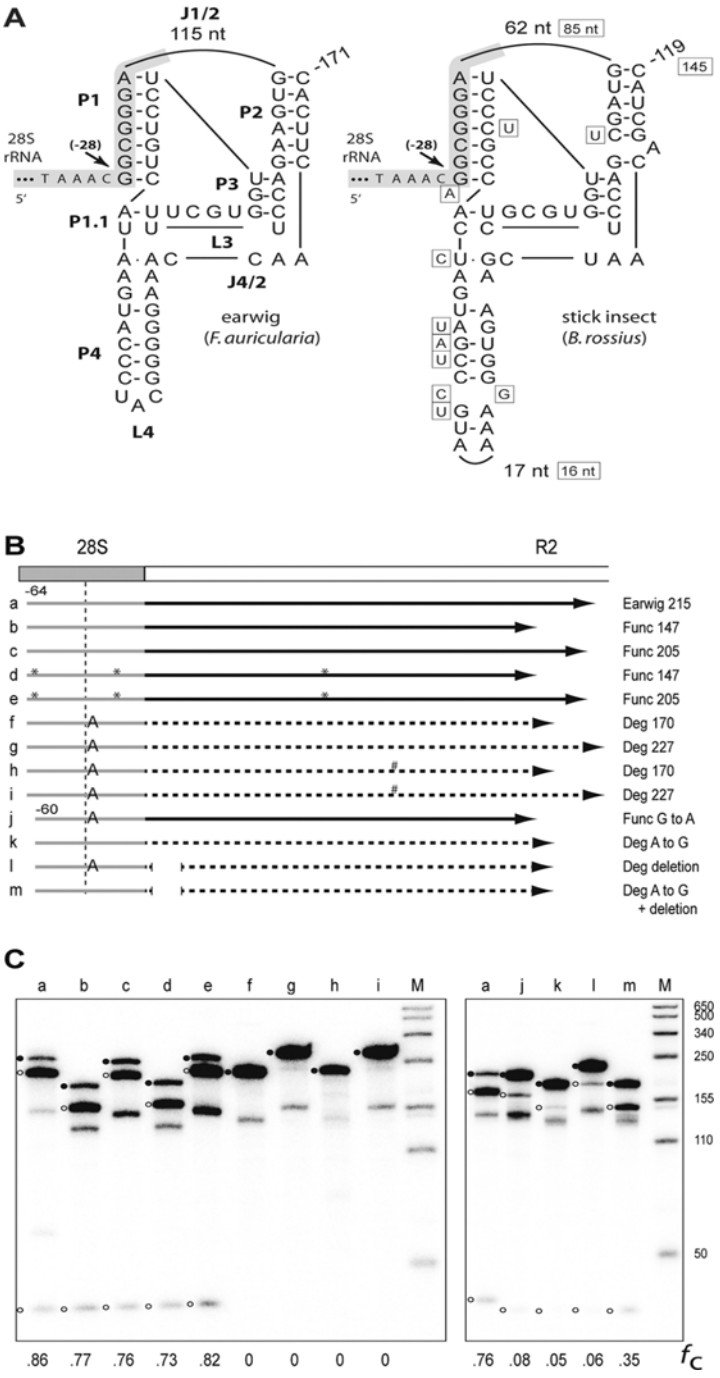


Fig 4. R2Br ribozyme structure and self-cleaving assay. (A) Structure of the R2 ribozyme from the earwig *F. auricularia* (left; [14, 15]) and that predicted from the *B. rossius* functional element (right) are presented. The predicted RNA secondary structure for the R2Br^{deg} element is similar to the latter with nucleotide differences indicated outside the functional element structure (boxed nucleotides). Only the number of R2 nucleotides in the J1/2 loop and the number of nucleotides in the L4 loop of *B. rossius* are shown. The 28S gene sequences within and upstream of the ribozymes are shaded gray. Arrows indicate the observed or predicted R2 self-cleavage sites relative to the 3' R2 insertion site. P, base-paired region; L, loop at end of a P region; J, nucleotides joining base-paired regions [13]. (B) Diagram of a generic 28S gene (gray box)/ R2 5' end (white box) junction is shown. Arrows labeled a through m represent the *in vitro* generated RNAs derived from the earwig and stick insect R2s and tested for self-cleavage. RNAs derived from R2Br^{fun} are represented by a solid arrow while RNAs corresponding to R2Br^{deg} are represented by a dashed arrow. The extent of each RNA relative to the R2 element is indicated on the right. RNAs a-i begin at position-64 relative

to the R2 insertion site whereas RNAs j-m begin at position-60. RNAs j-m contain engineered mutations to test their effect on self-cleavage (see text). Nucleotide differences between the two functional copies (*) and between the two degenerate copies (#) are indicated. The vertical dashed line demarcates the predicted cleavage site. (C) 5% denaturing acrylamide gels showing the cleavage products for the RNAs in the co-transcription/self-cleavage assays. The uncleaved RNA (solid circles) and cleavage products (open circles) are indicated. Lanes are labeled with the corresponding letter from panel B. The fraction RNA that self-cleaved is indicated at the bottom. Lane M, RNA length markers with sizes indicated.

doi:10.1371/journal.pone.0121831.g004

resulted in an increase in self-cleavage to 35% (Fig. 4C, lane m). Free energy predictions by an RNA folding program (rna.urmc.rochester.edu/RNAstructureWeb/index.html) for the 5' junctions of the degenerate RNAs were consistent with these experimental results: the P1 structure predicted for the original R2Br^{deg} RNA involved base pairing between the upstream 28S sequences and a region in the J1/2 loop while the prediction for the P1 stem of the doubly modified RNA was as shown in Fig. 4A.

Discussion

The co-occurrence of multiple R2 elements within the same genome is a well-known situation, with instances of three, or even four, widely divergent R2 elements documented [5, 6, 25, 26]. In *B. rossius*, unlike the incidences in *Nasonia vitripennis* or *Tribolium castaneum* for example, this co-presence involves a functional element and its degenerate paralog residing within the same genome. Quite interestingly, the two variants have co-existed for at least 5 Myrs, and the degenerate variant shows the signature of ongoing replication despite the lack of a functional coding region [12]. A previous sequencing survey indicated that R2Br^{fun} and R2Br^{deg} showed quite different distribution patterns, the former being found as the only resident within the *B. r. rossius* genome, the latter being present with or without the functional element within different *B. r. redtenbacheri* populations [12]. The present study based on an intensive survey of the 5' ends of R2, however, indicates a different scenario as the two variants have been found in both sub-species. The present survey also reveals a peculiar outcome: all sequenced 5'-truncated elements belong to the R2Br^{fun} variant and, hence, the R2Br^{deg} elements are only full-length. As previously described, a 5' end truncation occurs during an integration event when either the reverse transcriptase falls off before reaching the end of the RNA or the RNA template itself is degraded. The occurrence of 5' truncations is a characteristic outcome of the non-LTR retrotransposon integration mechanism [27]. However, retrotransposition without evidence of 5' truncations could still be possible, for example, if the R2Br^{deg} element generated RNA which was less prone to degradation.

We, therefore, addressed the structure of the 5' end of the R2Br^{deg} RNA. Typically, mature R2 RNAs are produced by self-cleavage from 28S/R2 co-transcripts through an HDV-like, autocatalytic ribozyme encoded at the 28S/R2 5' junctions [13, 14]. Our analysis demonstrates that the junction sequences of R2Br are able to form the secondary structure of HDV-like ribozymes and would self-cleave in the 28S gene 28 nucleotides upstream of the insertion site, in line with other R2 ribozymes analyzed in insects [15]. Both R2Br^{fun} and R2Br^{deg} showed very similar secondary structures, the latter exhibiting point mutations that maintained the ribozyme structure but with a slightly longer J1/2 loop. Self-cleavage assays demonstrate that R2Br^{fun} has high levels of activity while R2Br^{deg} has no detectable catalytic activity. In particular, the analysis presented here indicates that at least two specific differences can make the R2Br^{deg} ribozyme ineffective: the "G" to "A" substitution in the 28S gene that occurs in all sequenced degenerate element junctions, and the additional 23 bp in the R2Br^{deg} 5' UTR that appears to interfere with the formation of the P1 stem. Rendering the R2Br^{deg} templates more like R2Br^{fun} by introducing either an "A" to "G" substitution in the 28S gene or a partial J1/2 loop

deletion restores low levels of self-cleavage activity suggesting a stepwise loss of ribozyme functionality. Finally, even if R2Br^{deg} is able to self-cleave from the co-transcript at an extremely low level, it is difficult to explain how the "A" nucleotide at the 5' end of a processed degenerate RNA would be copied into the upstream DNA target sequences (as seen in all sequenced junctions) based on previous models of R2 5' integration [15].

As previously evidenced [12] and confirmed here by sequence analysis and Tajima's D statistics, R2Br^{deg} still replicates. Significantly negative Tajima's D may indicate purifying selection or a sudden explosion of sequence duplications: while the first can be ruled out when dealing with TEs, and particularly with degenerate ones, the latter suggests that R2Br^{deg} arose once in the evolution of *Bacillus rossius* and then dramatically increased its copy number. Therefore, there is a mechanism by which R2Br^{deg} duplicates despite its inability to encode a protein or to self-cleave and without generating the typical pattern of 5'-truncated copies.

One possibility is that R2Br^{deg} behaves as a non-autonomous element, exploiting the enzymatic machinery of the co-existing R2Br^{fun}, as observed for the R1/R2-derived SIDE elements [28]. Having lost its ability to self-cleave, any non-functional R2 RNA would presumably include at a minimum the sequence from the entire 5' half of the 28S gene. Even assuming this RNA is stable and escapes mechanisms of rRNA quality control [29], its structure could interfere with the TPRT reaction. For example, a portion of the R2 RNA 5' end is bound by an R2 protein that mediates the second nick on the target DNA during the integration process and performs the second strand synthesis [10, 30]: it is possible that the additional 28S portion at the 5' end inhibits the TPRT process. Finally, the additional sequences at the 5' end might make it is less likely that degraded transcripts would give rise to the 5' truncations monitored at the 28S/R2 junction; however, the tendency for the R2 protein to fall off during first strand synthesis would not be effected and truncations generated in this manner should still occur.

Interestingly, 5'-truncated copies, all belonging to the R2Br^{fun} variant, show significantly negative Tajima's D and more than one sequence for each 5'-truncated variant occur (Fig. 1): this suggests that once 5'-truncated variants are produced they may further duplicate. On the contrary, Tajima's D calculated on the full-length R2Br^{fun} dataset is not different from zero indicating a mutation-drift equilibrium. This is consistent with a stable retrotransposon population where multiple R2 copies duplicate (multiple source model) and others are eliminated through mechanisms of genomic turnover or by drift.

Multiple copies of the same R2 insertion have been scored in *Drosophila* genomes, even if at a frequency lower than that scored in the present analysis, and are thought to be the product of the duplication of the inserted 28S by molecular drive processes [17, 31]. We, therefore, suggest an alternative scenario explaining both the R2Br^{deg} and 5'-truncated copies duplication: data presented here are consistent with the hypothesis of duplication through the spread of the original 28S-inserted copy by means of recombination events that are responsible for the concerted evolution of the ribosomal locus (molecular drive, [18, 19]).

This scenario could have major implications for the evolution of TEs targeting a specific site in tandem repeats. As previously reported, R2 belongs to an ancestral clade of non-LTR elements mostly characterized by site-specificity within tandem repeats [8, 9, 32]. The advantages of this strategy can be summarized in four main points: i) insertion in tandem repeats should bring little damage to the host, as uninserted repeats would still be present, ii) tandem repeats such as rDNA will guarantee a transcriptionally active population of insertion sites with which these elements can be co-transcribed, iii) the effect on the host of unequal recombination between retrotransposons should be not different than that between the tandem repeats themselves, while random insertions would lead to harmful ectopic recombination, and iv) molecular drive continuously removes insertions, leading, in the long term, to the survival of only functional (= active) copies. This appears to be the case in many *Drosophila* species where

R2 has a replication rate that counteracts the recombinational effects, leading to relatively small R2 populations composed mostly of active elements ([32] and reference therein). The stable maintenance, inheritance and duplication of R2Br^{deg} and 5'-truncated copies by 28S molecular drive, though, is deeply inconsistent with the last point and clearly shows that degenerate element survival is possible.

Metazoan genomes have a great excess of rDNA repeats so that, although there are several non-LTR elements and/or at least one DNA element potentially inserting within them [19], there are still enough units to produce the rRNA necessary for the cell to function. R2 elements, for example, vary widely in term of lineage richness and copy number. A single R2 lineage may occupy 10%-45% of the rDNA units in a *Drosophila simulans* genome or 0.5%-5% of the units in the tadpole shrimp *Triops cancriformis* [33, 34]. Moreover, an rDNA array can support up to four R2 lineages, as observed in the sea squirt *Ciona intestinalis*, and up to three lineages have been retrieved in beetles, in the tadpole shrimp *Lepidurus couesii* and in the turtle *Mauremys reevesi* [5, 6, 26, 35]. Although in these instances R2 lineages are all inferred to be functional, based on sequence analysis, it is possible that not all of them are actively retrotransposing. Taking into account the transcription domain model of R2 epigenetic regulation [36], Luchetti and Mantovani [5] suggested that lineages can be silenced while restricted to transcriptionally inactive regions of the rDNA array and/or unleashed when rearranged by unequal crossing over. In this way, also thanks to molecular drive's homogenizing forces, silenced lineages could be maintained over evolutionary time. Based on the empirical data gathered in *Drosophila*, Zhou et al. [31] simulated a population model to explain the interplay between rDNA molecular drive and R2 activity in light of the transcription domain model. They showed that i) the transcriptionally active rDNA domain can be established at each generation in region(s) with no R2 occurrence, ii) that R2 is active only when there is no choice but to transcribe one or more R2-interrupted rDNA unit(s) (e.g. after a contraction of the rDNA units copy number) and iii) that recombination occurs mainly in the transcriptionally active domains. This latter observation would explain why they observed very few R2 insertions duplicated by crossing over.

We have no estimates of R2Br occupancy within the stick insect genome (i.e. the percentage of 28S genes interrupted by an R2Br insertion), but previous inheritance studies highlighted a considerable plasticity of the insertion profile with new insertions and eliminations detectable even in a single generation [12]. Beside active retrotransposition, this speaks in favor of a remarkable rDNA array turnover and indicates that, based on the Zhou et al. [26] model and considering the R2Br^{fun} Tajima's D indicating a duplication/elimination equilibrium, one or more copies of R2Br^{fun} are currently within the rDNA transcriptionally active domain. On the other hand, the high rate of duplication of 5'-truncated and R2Br^{deg} elements observed here clearly contrasts with the Zhou et al. [31] model.

The different behavior of the rDNA/R2 relationship in stick insects with respect to fruit flies can be explained by three, non-mutually exclusive mechanisms. First, in stick insects, recombination is not restricted to the transcriptionally active domain of an rDNA array but may evenly occur throughout the array. Zhou et al. [31] demonstrated that this would lead to duplications of the same insertion but would decrease the number of different R2 insertions. In this view, it is also to be noted that rDNA loci are differentially distributed in the genomes of fruit flies and stick insects, *Bacillus* stick insects having multiple rDNA loci mostly located on autosomes [37, 38] while *D. simulans* showing a single rDNA locus on the X chromosome. Therefore, it is possible that, in stick insects, different R2 variants can be located and/or present in different proportions on different rDNA arrays. Where transcription is never or seldom active, but concerted evolution still takes place [31], recombination would occur and in this case the copy number of the retrotranspositionally inactive R2Br^{deg} or 5'-deleted R2Br^{fun} copies will increase.

Second, functional, 5'-truncated and degenerate elements are all within a transcriptionally active rDNA array because either the stick insect rDNA transcription machinery doesn't have the ability to select a relatively small R2-free region or there is not an insert-free region for the cell to select. This would place the control of R2 activity and proliferation on the equilibrium between transposition and elimination of active copies by molecular drive, as also data on R2Br^{fun} already suggested, and eventually on selective pressures against exceedingly active lineages.

Third, mechanisms of genomic turnover, and therefore concerted evolution, are slower in *B. rossius* than in *Drosophila*. In this case, R2-inserted 28S would be less efficiently eliminated in the short term and, on average, persist longer in stick insects. Modeling R2 insertion inheritance in *B. rossius* bisexual populations, we showed that the elimination of R2-inserted 28S is mainly driven by selection rather than recombination, in line with a possible low efficiency of genomic turnover mechanisms [12].

Whatever the mechanisms regulating their spread, the present condition of R2Br^{fun} 5'-truncated copies and R2Br^{deg} allows them to effectively avoid elimination from the array(s). It appears, therefore, that inserting specifically into a tandem repeat array allows dead R2 (and similar TEs) copies to survive and expand their population even if the coding capacity and the possibility to exploit the enzymatic machinery of functional elements have been lost. Taking into account that R2Br^{deg} has been maintained for at least 5 Myr, such strategy appears successful. It would be interesting to speculate about how long a dead element can survive in this way and what could be the consequence of their presence within the genome. Many parameters should be calculated to answer these questions (for example, the element occupancy, the rDNA recombination frequency and the number of rDNA loci) but it is likely that a dead element could survive until its population experiences severe contraction, making it more vulnerable to drift. An interesting point of the coexistence of multiple R2 lineages is that they may recombine, potentially generating further lineages [26]. Here we showed that R2Br^{fun} elements can recombine with degenerate ones; the gene conversion detected here involved a fragment carrying the stop codon identified in the R2Br^{deg} [12] integrating within a functional copy. Although the functional elements that underwent recombination are 5'-truncated, thus unable to retrotranspose, this finding suggests that the possible recombination between R2Br^{fun} and R2Br^{deg} may inactivate active copies rather than produce new lineages. This may be part of a possible trade-off in maintaining a number of degenerate element copies, and it will be an intriguing issue to better investigate in further studies.

Supporting Information

S1 Table. List of the primers used for obtaining DNA templates for T7 co-transcription/cleavage reactions. Primers have been utilized to generate: i) the original templates (OT), ii) the "G" to "A" mutation in the functional element ribozyme (G>A^{fun}), iii) the "A" to "G" mutation in the degenerate element ribozyme (A>G^{deg}), iv) the partial J1/2 loop deletion in the degenerate element ribozyme (J1/2 $^{loop\ del}$), v) the partial J1/2 loop deletion plus "A to G" double mutation (J1/2 $^{loop\ del}$+A>G^{deg}). For J1/2 $^{loop\ del}$ and J1/2 $^{loop\ del}$+A>G^{deg} final products, a nested PCR was performed involving first the *BrLoopDel+BrDeg_227REV* pair, then *T7/28S(G to A)+BrDeg_170REV* and *T7/28S(A to G)+BrDeg_170REV*, respectively.
(DOCX)

Acknowledgments

Authors wish to thank Livia Bonandin for helping in stick insects rearing and fruitful discussion. Authors also wish to thank the Editor and two anonymous Reviewers whose criticisms and suggestions substantially improved the manuscript.

Author Contributions

Conceived and designed the experiments: FM DGE AL BM. Performed the experiments: FM DGE CS. Analyzed the data: DGE AL. Contributed reagents/materials/analysis tools: DGE BM. Wrote the paper: FM DGE AL BM.

References

1. Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 2012; 509: 7–15. doi: 10.1016/j.gene.2012.07.042 PMID: 22921893
2. Huang CRL, Burns KH, Boeke JD. Active transposition in genomes. Annu Rev Genet 2012; 46: 651–675. doi: 10.1146/annurev-genet-110711-155616 PMID: 23145912
3. Arkhipova IR, Rodriguez F. Genetic and epigenetic changes involving (retro)transposons in animal hybrids and polyploids. Cytogenet Genome Res 2013; 140: 295–311. doi: 10.1159/000352069 PMID: 23899811
4. Jamburuthugoda VK, Eickbush TH. Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. Nucleic Acids Res 2014; 42: 8405–8415. doi: 10.1093/nar/gku514 PMID: 24957604
5. Luchetti A, Mantovani B. Non-LTR R2 element evolutionary patterns: phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages. PLoS ONE 2013; 8: e57076. doi: 10.1371/journal.pone.0057076 PMID: 23451148
6. Kojima KK, Fujiwara H. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. Mol Biol Evol 2005; 22: 2157–2165. PMID: 16014872
7. Kojima KK, Jurka J. R2 non-LTR retrotransposons not inserted into ribosomal RNA genes. Repbase Reports 2013; 13: 1215–1217.
8. Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 1999; 16: 793–805. PMID: 10368957
9. Malik HS, Eickbush TH. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. Genetics 2000; 154: 193–203. PMID: 10628980
10. Christensen SM, Ye J, Eickbush TH. RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc Natl Acad Sci USA 2006; 103: 17602–17607. PMID: 17105809
11. Pérez-Gonzalez CE, Eickbush TH. Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans*. Genetics 2001; 158: 1557–1567. PMID: 11514447
12. Bonandin L, Scavariello C, Luchetti A, Mantovani B. Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta Phasmida). Insect Mol Biol 2014; 23: 808–820. doi: 10.1111/imb.12126 PMID: 25134735
13. Eickbush DG, Eickbush TH. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA co-transcript. Mol Cell Biol 2010; 30: 3142–3150. doi: 10.1128/MCB.00300-10 PMID: 20421411
14. Ruminski DJ, Webb C-HT, Riccitelli NJ, Lupták A. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. J Biol Chem 2011; 286: 41286–41295. doi: 10.1074/jbc.M111.297283 PMID: 21994949
15. Eickbush DG, Burke WD, Eickbush TH. Evolution of the R2 retrotransposon ribozyme and its self-cleavage site. PLoS ONE 2013; 8: e66441. doi: 10.1371/journal.pone.0066441 PMID: 24066021
16. Bibillo A, Eickbush TH. High processivity of the reverse transcriptase from a non-long terminal repeat retrotransposon. J Biol Chem 2002; 277: 34836–34845. PMID: 12101182
17. Stage DE, Eickbush TH. Origin of nascent lineages and the mechanisms used to prime second-strand DNA synthesis in the R1 and R2 retrotransposons of *Drosophila*. Genome Biol 2009; 10: R49. doi: 10.1186/gb-2009-10-5-r49 PMID: 19416522
18. Dover GA. Molecular drive: a cohesive mode of species evolution. Nature 1982; 299: 111–117. PMID: 7110332
19. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 2007; 175: 477–485. PMID: 17322354
20. Mantovani B, Passamonti M, Scali V. The mitochondrial cytochrome oxidase II gene in *Bacillus* stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. Mol Phylogenet Evol 2001; 19: 157–163. PMID: 11286500

21. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 2013; 30: 2725–2729. doi: 10.1093/molbev/mst197 PMID: 24132122

22. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009; 25: 1451–1452. doi: 10.1093/bioinformatics/btp187 PMID: 19346325

23. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 2012; 61: 539–542. doi: 10.1093/sysbio/sys029 PMID: 22357727

24. Bibillo A, Eickbush TH. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. J Biol Chem 2004; 279: 14945–14953. PMID: 14752111

25. Stage DE, Eickbush TH. Maintenance of multiple lineages of R1 and R2 retrotransposable elements in the ribosomal RNA gene loci of *Nasonia*. Insect Mol Biol 2010; 19: 37–48. doi: 10.1111/j.1365-2583.2009.00949.x PMID: 20167016

26. Luchetti A, Mingazzini V, Mantovani B. 28S junctions and chimeric elements of the rDNA targeting non-LTR retrotransposon R2 in crustacean living fossils (Branchiopoda, Notostraca). Genomics 2012; 100: 51–56. doi: 10.1016/j.ygeno.2012.04.005 PMID: 22564473

27. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993; 72: 595–605. PMID: 7679954

28. Eickbush DG, Eickbush TH. R2 and R2/R1 hybrid non-autonomous retrotransposons derive d from internal deletions of full-length elements. Mob DNA 2012; 3:10. doi: 10.1186/1759-8753-3-10 PMID: 22621441

29. Doma MK, Parker R. RNA quality control in eukaryotes. Cell 2007; 131:660–668. PMID: 18022361

30. Moss WN, Eickbush DG, Lopez MJ, Eickbush TH, Turner DH. The R2 retrotransposon RNA families. RNA Biol 2011; 8:714–718. doi: 10.4161/rna.8.5.16033 PMID: 21734471

31. Zhou J, Eickbush MT, Eickbush TH. A population genetic model for the maintenance of R2 retrotransposons in rRNA gene loci. PLoS Genet 2013; 9: e1003179. doi: 10.1371/journal.pgen.1003179 PMID: 23326244

32. Eickbush TH. R2 and related site-specific non-LTR Retrotransposons. In: Craig N, Craigie R, Gellert M, Lambowitz A, editors. Mobile DNA II. Washington D.C.: American Society of Microbiology Press; 2002. pp. 813–835.

33. Zhou J, Eickbush TH. The pattern of R2 retrotransposon activity in natural populations of *Drosophila simulans* reflects the dynamic nature of the rDNA locus. PLoS Genet 2009; 5: e1000386. doi: 10.1371/journal.pgen.1000386 PMID: 19229317

34. Mingazzini V, Luchetti A, Mantovani B. R2 dynamics in *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution. Heredity 2011; 106: 567–575. doi: 10.1038/hdy.2010.86 PMID: 20628416

35. Kojima KK, Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol 2004; 21: 207–217. PMID: 12949131

36. Eickbush DG, Ye J, Zhang X, Burke WD, Eickbush TH. Epigenetic regulation of retrotransposons within the nucleolus of *Drosophila*. Mol Cell Biol 2008; 28: 6452–6461. doi: 10.1128/MCB.01015-08 PMID: 18678644

37. Manaresi S, Marescalchi O, Scali V. Ag-detected NOR and C-banding patterns in *Bacillus rossius* (Insecta Phasmatodea) from Siciliy. Caryologia 1991; 44: 265–286.

38. Tinti F, Scali V. C-banding, Ag-NOR localization and chromosomal repatterning in Sardinian *Bacillus atticus* (Insecta, Phasmatodea). Boll Zool 1991; 58: 235–243.

Table S1. List of the primers used for obtaining DNA templates for T7 co-transcription/cleavage reactions. Primers have been utilized to generate: i) the original templates (OT), ii) the "G" to "A" mutation in the functional element ribozyme (G>A^{fun}), iii) the "A" to "G" mutation in the degenerate element ribozyme (A>G^{deg}), iv) the partial J1/2 loop deletion in the degenerate element ribozyme (J1/2 $^{loop\ del}$), v) the partial J1/2 loop deletion plus "A to G" double mutation (J1/2 $^{loop\ del}$+A>G^{deg}). For J1/2 $^{loop\ del}$ and J1/2 $^{loop\ del}$+A>G^{deg} final products, a nested PCR was performed involving first the *BrLoopDel+BrDeg_227REV* pair, then *T7/28S(G to A)+BrDeg_170REV* and *T7/28S(A to G)+BrDeg_170REV*, respectively.

Amplicon	Primer names	Upstream primer (5' > 3')	Downstream primer (5' > 3')
OT	*T7/28S(-63G)*	TAATACGACTCACTATAGGGgCAAAGTGAAGAAATTCAACGAAGCG	
OT	*T7/28S(-63T)*	TAATACGACTCACTATAGGGtCAAAGTGAAGAAATTCAACGAAGCG	
OT	*Br^{fun}_147REV*		CCAGCCCCGGGTACCAGC
OT	*Br^{fun}_205REV*		CCATCCGAGGCTTCTTTTTGTAAGAAC
OT	*Br^{deg}_170REV*		CTGGCCCCGGGTACCATC
OT	*Br^{deg}_227REV*		CATCCGAGGCTTCATTTTGCGAAATC
G>A^{fun}	*T7/28S(G to A)*	TAATACGACTCACTATAGGGAAGTGAAGAAATTCAACGAAGCGCGGGTAAACAGCGGGAG	
G>A^{fun}	*Br^{fun}_147REV*		CCAGCCCCGGGTACCAGC
A>G^{deg}	*T7/28S(A to G)*	TAATACGACTCACTATAGGGAAGTGAAGAAATTCAACGAAGCGCGGGTAAACGGCGGGAG	
A>G^{deg}	*Br^{deg}_170REV*		CTGGCCCCGGGTACCATC
J1/2 $^{loop\ del}$	*BrLoopDel*	GCGCGGGTAAACAGCGGGAGTAACTATGACTCTCTTGGAGGCTGAGGAGATCG	
J1/2 $^{loop\ del}$	*Br^{deg}_227REV*		CATCCGAGGCTTCATTTTGCGAAATC
J1/2 $^{loop\ del}$	*T7/28S(G to A)*	TAATACGACTCACTATAGGGAAGTGAAGAAATTCAACGAAGCGCGGGTAAACAGCGGGAG	

J1/2 $^{\text{loop del}}$	*Br^{deg}_170REV*		CTGGCCCCGGGTACCATC
J1/2 $^{\text{loop del}}$+A>G^{deg}	*BrLoopDel*	GCGCGGGTAAACAGCGGGAGTAACTATGACTCTCTTGGAGGCTGAGGAGATCG	
J1/2 $^{\text{loop del}}$+A>G^{deg}	*Br^{deg}_227REV*		CATCCGAGGCTTCATTTTGCGAAATC
J1/2 $^{\text{loop del}}$+A>G^{deg}	*T7/28S(A to G)*	TAATACGACTCACTATAGGGAAGTGAAGAAATTCAACGAAGCGCGGGTAAACGGCGGGAG	
J1/2 $^{\text{loop del}}$+A>G^{deg}	*Br^{deg}_170REV*		CTGGCCCCGGGTACCATC

2.1.3. First case of R2 non-LTR retrotransposon horizontal transfer

The R2 complete element in gonochoric populations of the strictly bisexual *B. grandii grandii, B. grandii benazzii and B. grandii maretimi* and in obligatory parthenogenetic population of *B. atticus atticus* were characterized, for a total of six R2 complete sequences. In the *B. grandii grandii* genome, two R2 elements occur. One of these probably represents a different R2 lineage: it shows in fact two Zinc Finger motifs, at variance of the usual one ZF motif of all other *Bacillus* R2 elements. Also *B. grandii maretimi* has two R2 elements, while *B. grandii benazzii* and *B. atticus atticus* have only one element. When compared with R2Br^{fun} and R2Br^{deg}, R2s from *B. g. maretimi* and *B. g. benazzii* are less divergent (3.8% - 5.2%) with respect to the R2 element from the *B. g. grandii* ^{A} and *B. a. atticus* (17.6% - 56.5%). The observed lower divergence among *B. rossius, B. g. maretimi* and *B. g. benazzii* R2 elements has led me to think of a possible case of horizontal transfer (HT). To infer a horizontal transfer event, usually, three approaches can be utilized (Schaack et al., 2010):

I) high sequence similarity of the TE from different host species exceeding the level that would be expected given the divergence time of the hosts

II) incongruence between TE and host phylogeny

III) patchy distribution of the TE within a group of taxa

All analyses point to possible HT events of the R2 element in the *Bacillus* genus.

This part of my work is presented in the paper:

Scavariello C., Luchetti A. and Mantovani B. First case of horizontal transmission of R2 non-LTR retrotransposon in *Bacillus* stick-insects genome (Insecta Phasmida). (In preparation)

The results were also presented at the following symposia/Congresses:

Scavariello C., Luchetti A. and Mantovani B. 2015. First case of R2 non-LTR retrotransposon horizontal transfer: data from *Bacillus* stick-insects genome (Insecta Phasmida). VI Congress of the Italian Society of Evolutionary Biology SIBE-ISEB Poster n°06 pag. 71 (Atti di convegno-abstract). Bologna, 31/08/2015 – 03/09/2015

First case of horizontal trasmission of R2 non-LTR retrotransposon in *Bacillus* stick-insects genome (Insecta Phasmida).

Claudia Scavariello, Andrea Luchetti and Barbara Mantovani

Abstract

Horizontal transfer (HT) is an event in which genetic material is transferred from one species to another, even distantly related. Previous studies on R1 and R2 elements, which were only conducted on partial RT domains sequences, showed that they have been vertically transmitted since the Radiata-Bilateria split (Eickbush & Eickbush, 1995; Malik et al., 1999; Kojima and Fujiwara 2005). Here we present the first possible case of R2 HT in the *Bacillus* genus (Phasmatodea). Analyses of sequence similarity conducted on R2 full-length elements, their “patchy” distribution and the results of divergence-versus-age analysis agree and support HT events.

Introduction

Transposable elements (TEs) are genetic elements that are able to replicate and move within many eukaryotic genomes. TEs can be divided into two classes on the basis of their transposition mechanisms. Class I elements, also known as retroelements, move by reverse transcription of an RNA intermediate; the main distinction that is made within Class I retroelements is based on the presence or the absence of long terminal repeats (LTRs) flanking the element body. Non-LTR TEs include long interspersed elements (LINEs) and non autonomous short interspersed elements (SINEs). Class II elements, or transposons, move predominantly via a DNA-mediated mechanism of excision and insertion, although a few transposons move through a rolling-circle mechanism.

TEs are typically vertically inherited, passing from parents to offspring with subsequent duplications. Growing evidences though suggest that TEs transmission may not be restricted to vertical inheritance but horizontal transfers between lineages may occur.

Horizontal transfer (HT) is the passage of genetic material between reproductively isolated, even distantly related species. It has been proposed as an essential part of the lifecycle of some types of TEs. TEs are, in fact, subject to host suppression mechanisms, limiting their mobility and their copy number expansion , and also to stochastic losses either by segregation. In this scenario, HT events can be considered as an opportunity for TEs to invade a new genome and persist in evolution(Silva et al.,2004).

HT is apparently more frequent among Class II than Class I TEs, and among Class I TEs it is more frequent for LTR retrotransposons. This is possibly linked first to the different transposition mechanisms used by the two classes. DNA transposons have a more stable double-stranded DNA intermediate, while retrotransposons have in general terms a relatively unstable RNA intermediate that is reverse-transcribed directly into the chromosomal target site, so that the possible transfer outside the cell nucleus is decreased. Yet, between Class I elements, there is a small group of LTR retrotransposons (as Gipsy elements) that are similar to retroviruses in terms of replication mechanism and structural organisation (Malik et al., 2000). These elements have two open reading frame (ORF), corresponding to retroviral *gag* and *pol* genes, and an extra ORF in the same position as the *env* gene found in retrovirus genomes. This ORF consists of a putative transmembrane domain at the C-terminus, multiple putative N-glycosylation sites, and putative protease cleavage sites, similar to the cleavage sites in a variety of retroviral *env* proteins, at conserved positions (Song et al.,1994). The protein encoded by this ORF is from a structural and functional point of view comparable to the envelope protein present in retroviruses. For this reason, these LTR retrotransposons could be able to move easily between species by infecting new cells (Havecker et al., 2004). Without virus-like envelope proteins, other TEs would require a vector to facilitate their HT. On the whole, DNA transposons and LTR retrotransposons are, therefore, more likely to be capable of HT (Schaack et al., 2010). Concerning non-LTR retrotransposons, several previous phylogenetic studies, based on the RT domain, have demonstrated that these retroelements can be divided into 11 clades. Four of these clades have been maintained in diverse eukaryotes, while the other clades are present in no more than two taxonomic groups each one. Divergence *vs* age analysis of these 11 clades did not find any reliable evidence for HTs for these elements during the past 600 Myr. (Eickbush & Eickbush, 1995; Malik et al., 1999). Extensive studies specifically carried out on the R1 and R2 non-LTR retrotransposons showed that these elements have been vertically transmitted since the Radiata-Bilateria split (Eickbush & Eickbush, 1995; Kojima and Fujiwara 2005).

However, recently, some new cases of HT concerning non-LTR retrotransposons have been found: CR1 clade in *Maculinea* butterflies and *Bombyx* (Novikova et al., 2007); L1 and BovB elements among the vertebrates (Ivancevic et al., 2013); Sauria SINE in snakes and rodents (Piskurek and Okada, 2007) and R4 clade between *Aedes* and *Anopheles* (Biedler et al., 2015).

Among non-LTR retroelements, R2 is the most investigated one. R2 has a specific

insertion into the 28S ribosomal genes, disrupting its functionality. It has a single open reading frame (ORF) flanked by two untranslated sequences of variable length. The ORF comprises the central reverse transcriptase (RT) domain, the DNA-binding motifs at the N-terminus and the restriction enzyme-like endonuclease (RLE) domain at the C-terminus. The C-terminal end of the R2 protein includes a cysteine-histidine (zinc finger, ZF) motif (CCHC), while the N terminal domain can contain one (CCHH), two (CCHH + CCHH or CCHC + CCHH), or three (CCHH + CCHC + CCHH) ZF motifs (Kojima & Fujiwara, 2005; Luchetti & Mantovani, 2013). R2 is subject to significant changes in copy number even within a single species (Jakubczak et al. 1992). These are due to its rapid turnover, with high rates of retrotransposition and elimination (Perez-Gonzalez and Eickbush 2001). R2, being embedded in the repeated rDNA loci, is also subject to molecular drive, the dual force involving variant repeat fixation (at the genome level) and homogenization (at the population level). This dynamics, based on genomic turnover mechanisms and bisexual reproduction, respectively, should lead to the rDNA loci concerted evolution, i.e tandemly repeated sequences are more similar to each other within than between reproductive units.

R2 elimination and diversification make it difficult to trace the its evolution. We recently analyzed R2 in the stick insect *Bacillus rossius* (R2Br). In addition to a canonical element encoding a 1054 amino acid sequence comprising all known R2 domains and a single ZF motif (CCHH type) at the N-terminal end (named R2Br^{fun}), the genome of this facultative parthenogenetic species hosts a degenerate but closely related element. R2Br^{deg} exhibits 14 frameshift mutations and one stop codon within the open reading frame (Bonandin et al, 2014).

In this paper we isolate and characterize the R2 elements present in the other Italian species of the *Bacillus* genus (the obligatory parthenogenetic *B. atticus* and the gonochoric *B. grandii*) with the aim also to verify possible HT events.

To infer on the possibility of R2 HT events in the *Bacillus* genus, we used the three traditional criteria (Silva et al.,2004; Loreto et al., 2008):

I) TE from different host species with sequence similarity higher than that expected on the basis of the hosts divergence;
II) incongruences between TE and host phylogeny;
III) patchy distribution of the TE within a group of taxa.

Materials and Methods

Sampling, DNA isolation and R2 sequencing

Stick insects samples of the obligatory parthenogenetic *Bacillus atticus* and of the subspecies of the strictly gonochoric *Bacillus grandii* were field-collected in Sicily and either immediately frozen at -80 °C. Full-length R2 elements were isolated and characterized from one female of *Bacillus atticus* (Scoglitti, Sicily; BattSCO♀25) and from one male of each *B. grandii* subspecies: *B. grandii grandii* (Ponte Manghisi; BggPMA♂54), *B. grandii benazzii* (Torre Bennistra; BgbTBE♂4) and *B. grandii maretimi* (Marettimo Island, Sicily; BgmMAR♂ 2).

Total DNA was extracted from single stick insect legs with the standard phenol/chloroform protocol. R2 was isolated through PCR amplification, cloning and sequencing as reported in Bonandin et al. (2014). Universal and specifically designed primers used in this study are reported in Table 1 and Fig. 1

Sequence analysis

Sequences were edited and assembled using MEGA v. 6.0 (Tamura et al. 2013) and open reading frames (ORF) were searched with the ORF Finder tool server (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Nucleotide and amino acid sequence alignments has been carried out using MAFFT 7.2 (Katoh and Stadley, 2013) with L-INS-i parameters. Sequence divergences, calculated as uncorrected *p* distances, and Maximum Likelihood phylogenetic analyses, with 500 bootstrap replicates for nodal support, have been carried out using MEGA v. 6.0. Two phylogenetic analyses have been performed. The first analysis was based on inferred amino acid sequences encompassing the reverse transcriptase (RT) and the restriction-like enzyme domains and included, for appropriate comparisons, selected R2 elements from several metazoan (Table S1). The second analysis included only *Bacillus* elements (with the exception of $R2Bg^{B}$) and was based on nucleotide sequences. A divergence versus age analysis has been performed by plotting amino acid sequence divergences of the RT domain against the host age split. We added to the analysis a number of vertically inherited elements to have a comparison with possible instances of HT (Table S2): this is based on the principle that elements deriving from a HT are less divergent than the expected from the host-split age, while paralogous

lineages are more divergent then expected (Malik et al., 1999; Biedler et al., 2015). We used studentized deleted residual values (RStudent) in order to identify comparisons that have major influence on the best-fit line and that, therefore, can be outlier; RStudent absolute values exceeding 2 are considered as such (Biedler et al. 2015).

Results

R2 sequence characterization

Seven new R2 elements have been isolated from the genome of five *Bacillus* stick insect species; their sequence structures were depicted in figure 1.

One R2 element was found in *B. atticus atticus* (hence called R2Ba); it resulted 3507 bp long, excluding the 3' terminal poly-(A) tail of 9 nucleotides, with an ORF of 3177 bp encoding a protein of 1058 amino acid.

Two R2 elements were isolated from *B. grandii grandii*. The first ($R2Bg^{A}$) is 3513 bp long, excluding the poly-(A) tail of 7 nucleotides and resulted having two overlapped ORFs of 1545 bp and 1698 bp. The analysis of the proteins suggests the they are, actually, part of 3243 bp long ORF that encodes a protein of 1079 amino acid in which a frame shift mutation occurred (at pos. 1726). The same can be observed for the second *B. g. grandii* element ($R2Bg^{B}$), that is 4832 bp long, excluding the poly-(A) tail, with a putative ORF of 4264 bp in which a frameshift mutation occurred (at pos. 2191). Moreover, two stop codons (pos. 2913 and pos. 4722) also are recorded. This element also shows three sequence duplications 333 bp, 68 bp and 57 bp long, respectively; the first and second duplication are within the ORF while last one is in 5' UTR.

The element found in *B. grandii benazzii* (R2Bb) is 2754 bp long, excluding the poly-(A) tail, with an ORF of 2244 bp encoding a protein of 747 amino acid.

In the genome of *B. grandii maretimi*, two R2 elements were retrieved. The first (R2Bm) is 3485 bp long, excluding the poly-(A) tail. Its ORF appears of 3164 bp long but, again, it has one frame shift mutation disrupting the sequence at 875 bp. The second element is 3059 bp long, excluding the poly-(A) tail and it is nearly identical to R2Bm (0.4% of divergence), with the exception of a 426 bp deletion located between positions 1035 - 1460. It will be, therefore, called $R2Bm^{del}$. This element shows a complete ORF of 2739 bp encoding a protein of 912 amino acid.

Interestingly an element nearly identical to R2Bm^{del}, showing the 99% of similarity and having the same deletion with respect to R2Bm, has been retrieved in a 5' end sequence survey carried out in *B. rossius* (Martoni et al., 2015). It will be, thus, referred as R2Br^{del}. This R2 fragment, therefore, encompass the whole 5' UTR and the first 897 bp of the ORF; the protein sequence obtained from that part of the ORF is of 121 amino acids, and the RT domain was not covered.

All isolated elements showed one zinc finger motif, CCHH-type, at the protein N-terminal end, except R2Bg^{B} that showed two zinc finger motifs, CCHC+CCHH-type (Luchetti and Mantovani, 2013).

The R2Bg^{B} element is the most divergent at the nucleotide level, showing divergence ranging from 55.9% (R2Bg^{A}) to 57.1% (R2Bm^{del}; Table 2). When compared with *B. rossius* elements, R2Br^{fun} and R2Br^{deg} (Bonandin et al., 2014), R2s from *B. g. maretimi* and *B. g. benazzii* are less divergent (3.8% - 5.2%) than to R2 elements from the *B. g. grandii* and *B. a. atticus* (17.6% - 56.5%: Table 2). Moreover, R2Br^{del} divergence ranges from 0.4% (R2Bm) to 56.9% (R2Bg^{B}; Table 2).

Phylogenetic analyses

The Maximum Likelihood tree computed on RT+RLE amino acid sequences (figure 2) clearly reflects the clustering pattern based on the number of zinc-finger motif at the N-terminal end (Kojima and Fujiwara 2005; Luchetti and Mantovani, 2013). Accordingly to the amino acidic sequence structure, the element R2Bg^{B} cluster within clade B while the remaining *Bacillus* R2s fall within the clade D and are all grouped in a fully supported, monophyletic clade. Since amino acidic sequence is not available for the degenerate element R2Br^{deg} and only the N-terminal end is available for R2Br^{del}, we built another Maximum Likelihood tree based on the nucleotide sequence of *Bacillus* elements only, excluding R2Bg^{B} for its sequence is way too divergent and do not clearly share ancestry with the other elements. The tree overlap the topology of the *Bacillus* cluster in the amino acid sequence analysis (figure 3A), with the elements from *B. g. grandii* (R2Bg^{A}) and *B. a. atticus* showing a closer relationship and those from *B. rossius*, *B. g. benazzii* and *B. g. maretimi* included in the same clade. This is only partially congruent with the host species phylogeny (figure 3B), as *B. g. benazzii* and *B. g. maretimi* form a monophyletic clade with a sister relationship with that of *B. g. grandii* and *B. a. atticus* (Mantovani et al., 2001).

Divergence vs. age analysis

The analysis of RT amino acid divergence of selected non-LTR elements *vs* their relative host age revealed a significant correlation (Figure 4). Studentized deleted residuals with absolute values greater than 2 where obtained for the comparisons between *B. rossius* element R2Br^{fun} and the elements isolated in *B. g. benazzii* (R2Bb; RStudent = 2.30) and *B. g. maretimi* (R2Bm, RStudent = 2.20; R2Bm^{del}, Rstudent = 2,23). These values are, therefore, outliers well below the trendline, meaning that their divergence is significantly less than expected. This suggests that R2Bb, R2Bm and R2Bm^{del} could be the outcome of horizontal transfer.

Discussion

Non-LTR elements R2 are probably among the most widely studied TEs in both model and non-model organisms. One of the main issues regarding the R2 evolution is the frequent incongruence between its phylogeny and that of the host species, for which different hypotheses have been put forward. Kojima and Fujiwara (2005) analyzing a number of elements sampled across metazoan observed numerous instances of "local" congruences between R2 and host phylogenies, and did not identify cases of putative HT. Therefore, they concluded that the evolution of R2 is characterized by vertical inheritance with extensive paralogous lineages extinction/diversification. Luchetti and Mantovani (2013), in addition, suggested that in some ancestral genomes a library of unrelated elements that differentially amplified in the derived taxa, irrespective of their phylogeny might have been present. This would also explain why in some genomes are present multiple, unrelated R2 lineages.

Data reported here, though, suggest the possibility of HT between congeneric species of stick insects. In order to verify this aspect, we check if data fit three criteria that are considered relevant for validating a HT event: i) a divergence between elements from different species lower than expected on the basis of host split ages; ii) the phylogenetic incongruence between the putatively transferred element and the host species; iii) the patchy distribution among a group of taxa.

The phylogenetic analyses conducted on both amino acid and nucleotide sequences are well overlapping and indicate that, beside R2Bg^{B} that fall into the clade B, all *Bacillus* elements belong to the clade D and they are all included in a monophyletic cluster. Moreover, within this cluster, their phylogenetic relationships are not fully congruent with

the host phylogeny. In particular, elements from the species *B. g. maretimi* and *B. g. benazzii* cluster with *B. rossius* elements.

R2 elements of *B. grandii benazzii* and *B. grandii maretimi*, in fact, diverge of about 4%-5% from the functional variant of *B. rossius* R2, R2Br^{fun}, and of about 18% from the conspecific *B. grandii grandii* element R2Bg^{A}. Accordingly, the divergence versus age analysis indicate the comparisons R2Bb/R2Br^{fun} and R2Bm/R2Br^{fun} as significantly less divergent than expected on the basis of host split age (22.9 Myr; Mantovani et al., 2001). On the whole, therefore, these evidences suggests that R2 elements in *Bacillus* stick insects are vertically inherited but also that HT occurred between *B. rossius* and the *B. g. benazzii/B. g. maretimi* clade. In particular, two HT events took place: one involving the R2Bm^{del} element and one (or two) involving R2Bb-R2Bm element(s).

Patchy distribution refers to the presence of a given element in one lineage and its absence within the sister lineage. In the case of R2, beside the number of zinc fingers, there are no criteria to establish which lineage is which; on the other hand, Stage and Eickbush (2009), in *Drosophila* species, considered that element diverging >1% would represent separated lineages. If we take into account this threshold, then R2Bm^{del} and R2Br^{del} belong to the same lineage (but not to R2Bm due to the large deletion occurring within the ORF), which is absent in *B. grandii grandii* and *B. atticus*. Therefore, the "R2Bm-^{del}" lineage shows a patchy distribution. There are many instances of HT involving transposable elements (Schaack et al., 2010) and they can move laterally by means of vectors able to survive outside the host cell (for example virus; Piskurek and Okada 2011; Coates, 2015) or thanks to host organisms relationships, such as host-parasite (Gilbert et al., 2010) or predator-prey (Tang et al., 2015) ones. Actually we have not such evidences that may explain the HT between the two stick insects species, although a parasite-mediated transfer can be hypothesized: *Bacillus* species, in fact, have been recently observed infected by nematodes (Mantovani, personal observation). Interestingly, however, *B. rossius* and *B. g. benazzii* are the parental species of the hybridogenetic strain *B. rossius - grandi benazzii* (Scali et al., 2003). Here, chromosomes of the two species come into contact for one generation before to reconstitute the parental species, either *B. rossius* or *B. g. benazzii* depending to which species individual they backcross. Therefore, during these matings the two taxa exchanged the R2 even after the specie split (occurred about 6 Myr; Mantovani et al., 2000) thus maintaining a closer sequence similarity. This would, however, explain the HT observed for the R2Br^{fun}/R2Bb-R2Bm elements but not that suggested for the R2Bm^{del} element. In fact, hybridization between *B. rossius* and *B. g. maretimi* has never been observed. Although not ruling out other

possibilities (viruses, parasite-mediated exchange) it is possible to suggest that a way linked to the hybridization. *B. g. benazzii* and *B. g. maretimi* species split recently, ~6 Myr ago, thus it is possible that R2Bm^{del} element was already present in the ancestral genome. If hybridization occurred also between *B. rossius* and the ancestor of *B. g. benazzi*/*B. g. maretimi*, then R2Bm^{del} may have been transferred in the same way as R2Bb-R2Bm elements. We were not able to detect the presence of R2Bm^{del} in *B. g. benazzii,* thus we can hypothesize that in this species the element got extinct. On the whole, analyses presented here showed the existence of HTs involving R2 elements, a process previously ruled out for this retrotransposon on the basis of a metazoan-wide analysis (Kojima and Fujiwara, 2005). The peculiar way this element may have been transferred, i.e. by means of species hybridization, open an interesting question about the occurrence of *Bacillus* stick insect mobile genetic elements, taking into account the complex history of hybridization and hybridogenesis that characterize this genus.

References

BiedlerJ.K., Chen X. and Tu Z. (2015) Horizontal transmission of an R4 clade non-long terminal repeat retrotransposon between the divergent *Aedes* and *Anopheles* mosquito genera. Insect Molecular Biology 24(3), 331–337.

Bonandin, L., Scavariello, C., Luchetti, A. and Mantovani, B. (2014) Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta, Phasmida). Insect Molecular Biology 23(6): 808-820.

Burke WD, Eickbush DG,. Xiong Y, Jakubczak J, Eickbush TH (1993) Sequence relationship of retrotransposable elements *Rl* and *R2* within and between divergent insect species. Mol Biol Evol 10: 163-185.

Burke WD, Malik HS, Lathe WC 3, Eickbush TH (1998) Are retrotransposons long-term hitchhikers? Nature 392: 141–142.

Burke WD, Malik HS, Jones JP, Eickbush TH (1999) The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol Biol Evol 16: 502-511.

Coates B.S. (2015) Horizontal transfer of a non-autonomous Helitron among insect and viral genomes. BMC Genomics. 27;16:137

Dupeyron M., Leclercq S., Cerveau N., Bouchon D. and Gilbert C. (2014) Horizontal transfer of transposons between and within crustaceans and insects. Mob DNA. 2014 Jan 29;5(1):4. doi: 10.1186/1759-8753-5-4.

Eickbush, D.G. and Eickbush, T.H. (1995) Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. Genetics 139: 671–684.

Eickbush, T. and Malik, H. (2002) Origins and evolution of retrotransposons. In Mobile DNA II (Craig, N.L., Gellert, R.C.M. and Lambowitz, A.M., eds), pp. 1111–1144. American Society for Microbiology Press, Washington, DC.

Ghesini, S., Luchetti, A., Marini, M. and Mantovani, M. (2011) The non-LTR

retrotransposon R2 in termites (Insecta, Isoptera): characterization and dynamics. J Mol Evol 72: 296–305.

Gilbert C., Schaack S., Pace J.K. 2nd, Brindley P.J., Feschotte C. (2010) A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature. 2010;464:1347–1350.

Havecker ER, Gao X, Voytas DF. (2004) The diversity of LTR retrotransposons. Genome Biol 5: 225

Ivancevic AM, Walsh AM, Kortschak RD, Adelson DL. (2013) Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. Bioessays. 35(12):1071-82. doi: 10.1002/bies.201300072

Jakubczak JL, Zenni MK, Woodruff RC, Eickbush TH (1992) Turnover of R1 (type I) and R2 (type II) retrotransposable elements in the ribosomal DNA of *Drosophila melanogaster*. Genetics

Katoh, K. and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30: 772-780.

Kojima KK, Fujiwara H (2005) Long-term inheritance of the 28S rDNA-specific retrotransposon R2. Mol Biol Evol 22: 2157-2165.

Lathe, W.C., 3rd and Eickbush, T.H. (1997) A single lineage of r2 retrotransposable elements is an active, evolutionarily stable component of the *Drosophila* rDNA locus. Mol Biol Evol 14: 1232–1241

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Loreto ELS, Carareto CM and Capy P. (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity. 100:545–554.

Luchetti A, Mingazzini V, Mantovani B (2012) 28S junctions and chimeric elements of the rDNA targeting non LTR retrotransposon R2 in crustacean living fossils (Brachiopoda, Notostraca). Genomics 100: 51 – 56.

Luchetti A., Mantovani B. (2013) Non-LTR R2 element evolutionary patterns:

phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages. Plos One. 2013;8(2):e57076. doi: 10.1371/journal.pone.0057076.

Malik H.S., Burke W.D. and Eickbush T.H. (1999) The Age and Evolution of Non-LTR Retrotransposable Elements. Mol. Biol. Evol. 16(6):793–805.

Malik HS, Henikoff S.and Eickbush T.H. (2000) Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10:1307–1318.

Mantovani B., Scali V (1993) Genetic characterization of *Bacillus grandii maretimi* (n. subsp.) (INSECTA PHASMATODEA) in relation to its ecology. Vie Milieu, 43 (4) : 241-246

Mantovani B., Tinti F., Bachmann L., 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 parthenogenic taxa. Mol Biol Evol 14(12):1197-205.

Mantovani B., Passamonti M., Scali V. (1999) Genomic evolution in parental and hybrid taxa of the genus *Bacillus* (Insecta Phasmatodea). Ital. J. Zool., 66: 265-272.

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: 157–163.

Mingazzini, V., Luchetti, A. and Mantovani, B. (2011) R2 dynamics in *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution. Heredity 106: 567–575.

Novikova O.,Śliwińska E. , Fet V. , Settele J., Blinov A. and Woyciechowski M. (2007) CR1 clade of non-LTR retrotransposons from *Maculinea* butterflies (Lepidoptera: Lycaenidae): evidence for recent horizontal transmission. BMC Evol. Biol. 7:93.

Pérez-Gonzalez, C.E. and Eickbush, T.H. (2001) Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans*. *Genetics* 158: 1557–1567.

Piskurek O. and Okada N. (2007) Poxviruses as possible vectors for horizontal transfer of retrotransposons from reptiles to mammals. Proc Natl Acad Sci U S A. 104(29):12046-51.

Piskurek O. and Okada N. (2011) Tracking the ancestry of a deeply conserved eumetazoan SINE domain. Mol. Biol. Evol. 28(10):2727-30.

Scali, V., Passamonti, M., Marescalchi, O. and Mantovani, B. (2003) Linkage between sexual and asexual lineages: genome evolution in *Bacillus* sticks insects. Biol J Linn Soc 79: 137-50.

Schaack S, Gilbert C, Feschotte C. (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25: 537–46.

Silva JC, Loreto EL and Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol. 6:57–71.

Song SU, Gerasimova T., Kurkulos M., Boeke J.D. and Corces V.C. (1994) An env-like protein encoded by a *Drosophila* retroelement: Evidence that Gypsy is an infectious retrovirus. Genes Dev.; 8:2046–2057.

Tamura, K., Stecher, G, Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30(12):2725-9.

Tang Z., Zhang H.H., Huang K., Zhang X.G., Han M.J., Zhang Z. (2015) Repeated horizontal transfers of four DNA transposons in invertebrates and bats. Mob DNA. 17;6(1):3.

Figure legends

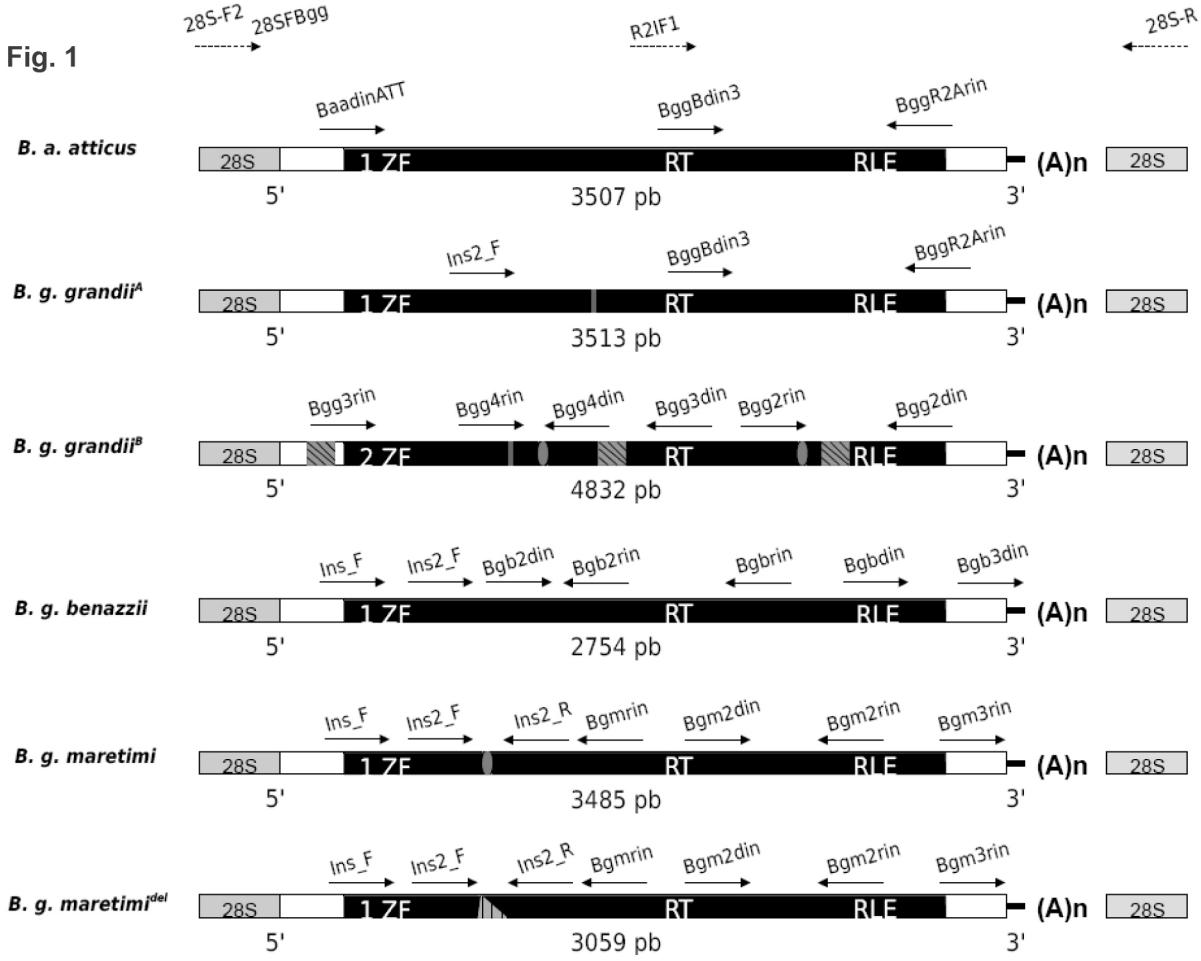
Fig.1: Schematic representation of R2 elements. Grey boxes indicate flanking 28S gene sequences; black boxes indicate the open reading frame (ORF) with the zinc finger (ZF), reverse transcriptase (RT) and restriction enzyme-like endonuclease (RLE) domains. The vertical grey lines indicate the frameshift mutations; the ovals represent the stop codon; the squares represent the duplications and the triangle represents the deletion.
Above each element, the primers used for sequencing are reported: the primers shared to all elements are indicate with dotted arrows.

Fig.2: Maximum Likelihood tree (-ln L= 35666.30) built on RT amino acid sequence of *Bacillus* elements and of elements representative of the four main clades (R2-A, R2-B, R2-C and R2-D) (listed in Table S1). Acronyms are as in Table S1; numbers at nodes represent bootstrap values ≥50%.

Fig.3: **A** Maximum Likelihood tree (ML; LogL= -2449,07) based on nucleotide sequences full-length R2 elements, except R2 Bgg^{B} due to its great nucleotide divergence; numbers at nodes represent bootstrap values ≥50% obtained after 500 replicates. **B** Maximum Likelihood tree based on mitochondrial COII gene of the genus *Bacillus* (Mantovani et al., 2001).

Fig.4: Points 1–9 represent non-LTR retrotransposons sequences arthropod comparisons taken from Malik et al. (1999). Points 10-24 represent non-LTR and R4 retrotransposons taken from Biedler et al. (2015). Points 25-29 are CRI non-LTR retrotransposons and points 30-34 are Jockey non-LTR retrotransposons taken from Novikova et al. (2007). Points a-l represent our R2 elements of each species of *Bacillus genus*. f, g and l are R2 elements of *B. grandii maretimi* and *B. grandii benazzii* a Rstudent values of 2,20, 2,23 and 2,30, respectively. For all elements was used the RT domain sequence. Statistics were calculated using log host age *vs*. amino acid divergence. **A** Plot host age *vs*. amino acid divergence with a logarithmic trendline. **B** Plot log host age *vs*. amino acid divergence with a linear trendline.

Fig. 1


28S-F2
28SFBgg
R2IF1
28S-R
B. a. atticus
BaadinATT
BggBdin3
BggR2Arin
28S
1 ZF
RT
RLE
(A)n
28S
5'
3507 pb
3'
B. g. grandii^A
Ins2_F
BggBdin3
BggR2Arin
28S
1 ZF
RT
RLE
(A)n
28S
5'
3513 pb
3'
B. g. grandii^B
Bgg3rin
Bgg4rin
Bgg4din
Bgg3din
Bgg2rin
Bgg2din
28S
2 ZF
RT
RLE
(A)n
28S
5'
4832 pb
3'
B. g. benazzii
Ins_F
Ins2_F
Bgb2din
Bgb2rin
Bgbrin
Bgbdin
Bgb3din
28S
1 ZF
RT
RLE
(A)n
28S
5'
2754 pb
3'
B. g. maretimi
Ins_F
Ins2_F
Ins2_R
Bgmrin
Bgm2din
Bgm2rin
Bgm3rin
28S
1 ZF
RT
RLE
(A)n
28S
5'
3485 pb
3'
B. g. maretimi^del
Ins_F
Ins2_F
Ins2_R
Bgmrin
Bgm2din
Bgm2rin
Bgm3rin
28S
1 ZF
RT
RLE
(A)n
28S
5'
3059 pb
3'

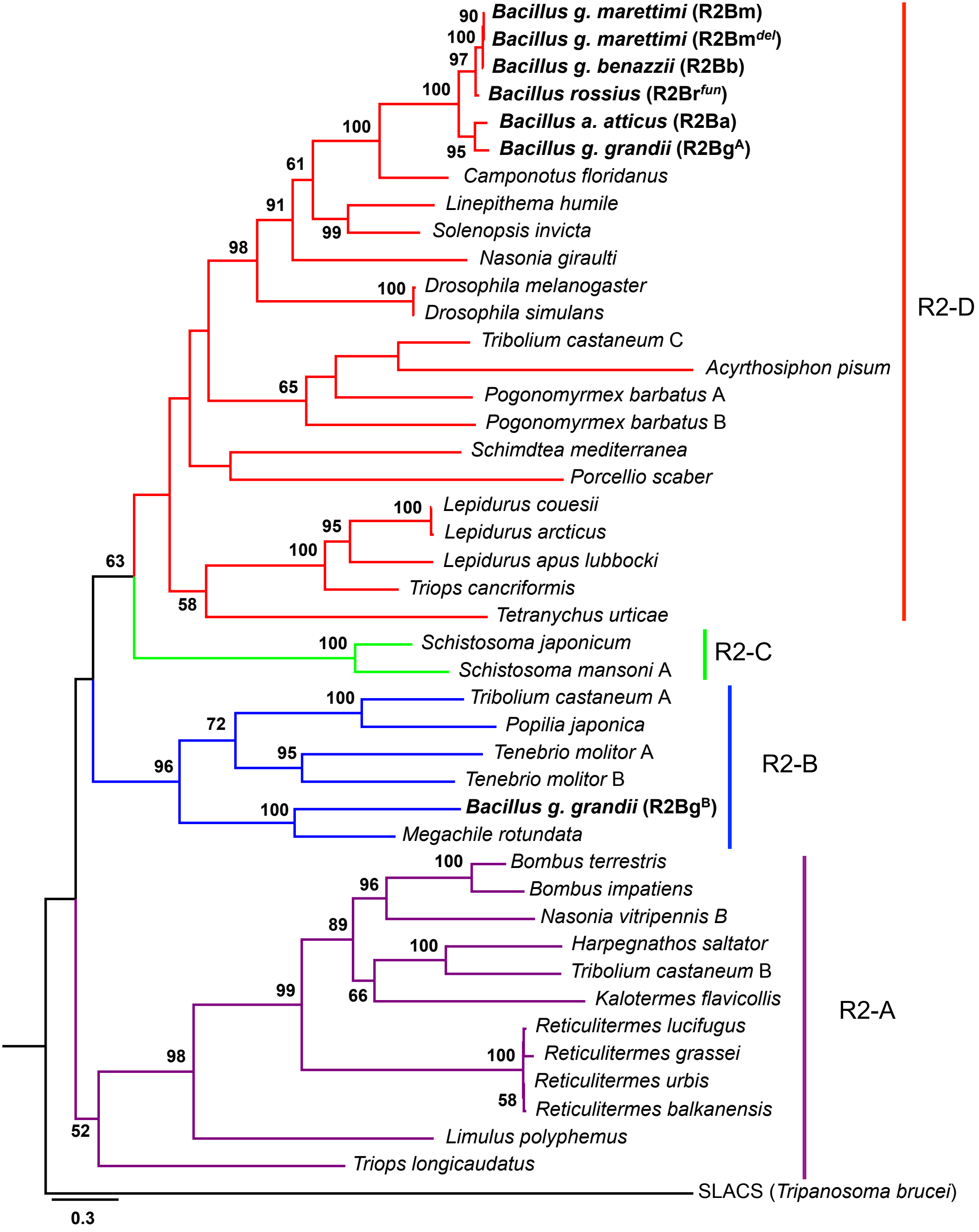
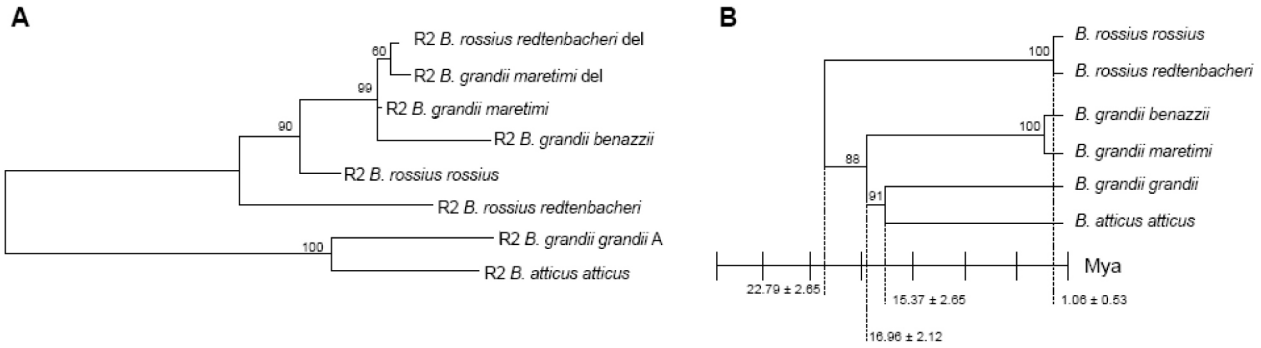

Bacillus g. marettimi (R2Bm)
Bacillus g. marettimi (R2Bm^del)
Bacillus g. benazzii (R2Bb)
Bacillus rossius (R2Br^fun)
Bacillus a. atticus (R2Ba)
Bacillus g. grandii (R2Bg^A)
Camponotus floridanus
Linepithema humile
Solenopsis invicta
Nasonia giraulti
Drosophila melanogaster
Drosophila simulans
Tribolium castaneum C
Acyrthosiphon pisum
Pogonomyrmex barbatus A
Pogonomyrmex barbatus B
Schimdtea mediterranea
Porcellio scaber
Lepidurus couesii
Lepidurus arcticus
Lepidurus apus lubbocki
Triops cancriformis
Tetranychus urticae
Schistosoma japonicum
Schistosoma mansoni A
Tribolium castaneum A
Popilia japonica
Tenebrio molitor A
Tenebrio molitor B
Bacillus g. grandii (R2Bg^B)
Megachile rotundata
Bombus terrestris
Bombus impatiens
Nasonia vitripennis B
Harpegnathos saltator
Tribolium castaneum B
Kalotermes flavicollis
Reticulitermes lucifugus
Reticulitermes grassei
Reticulitermes urbis
Reticulitermes balkanensis
Limulus polyphemus
Triops longicaudatus
SLACS (Tripanosoma brucei)
R2-D
R2-C
R2-B
R2-A
0.3

Fig. 3


A
R2 B. rossius redtenbacheri del
R2 B. grandii maretimi del
R2 B. grandii maretimi
R2 B. grandii benazzii
R2 B. rossius rossius
R2 B. rossius redtenbacheri
R2 B. grandii grandii A
R2 B. atticus atticus
60
99
90
100
B
B. rossius rossius
B. rossius redtenbacheri
B. grandii benazzii
B. grandii maretimi
B. grandii grandii
B. atticus atticus
100
100
88
91
Mya
22.79 ± 2.65
16.96 ± 2.12
15.37 ± 2.65
1.06 ± 0.53

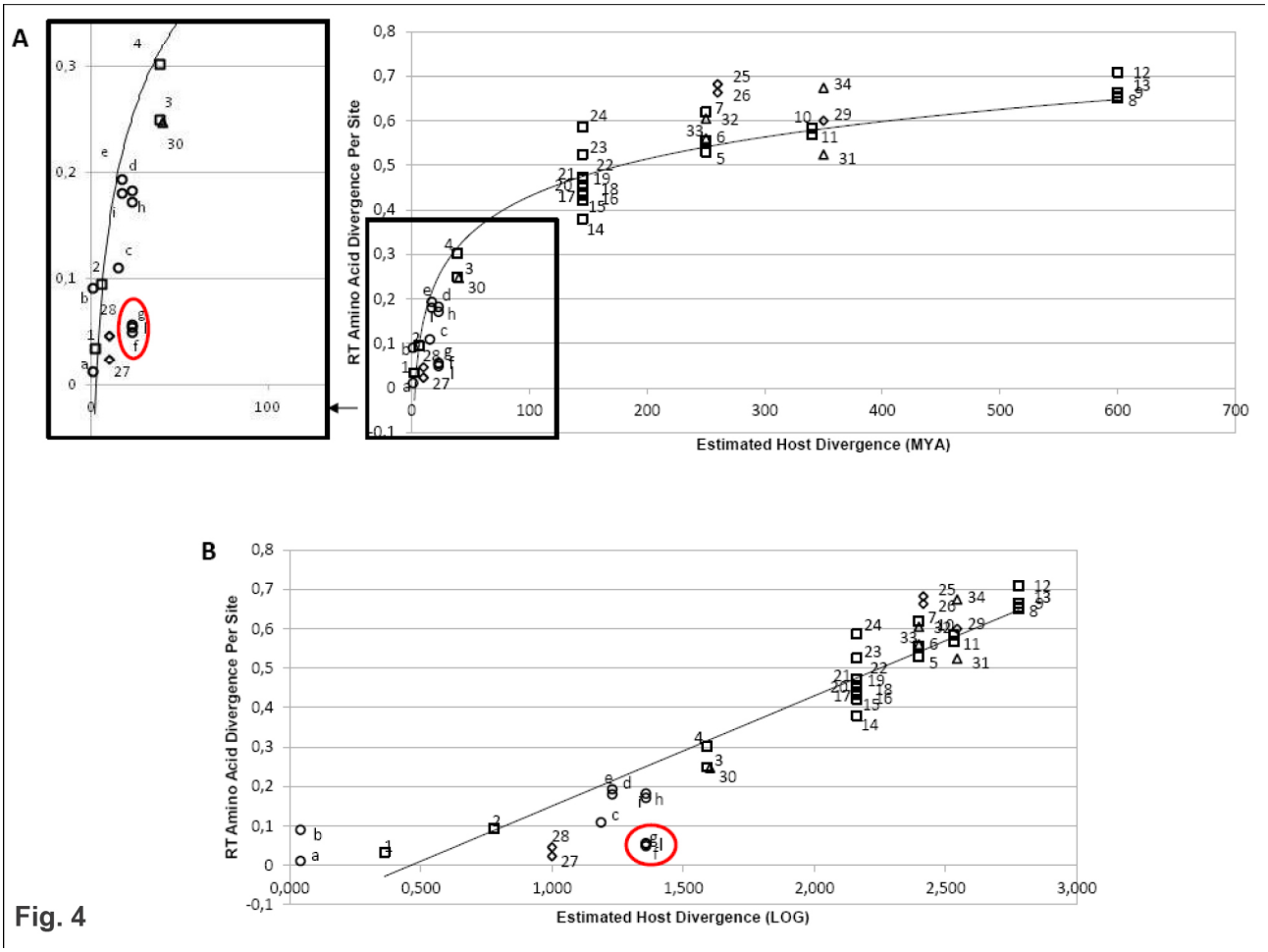

A
RT Amino Acid Divergence Per Site
Estimated Host Divergence (MYA)
B
RT Amino Acid Divergence Per Site
Estimated Host Divergence (LOG)

Fig. 4

Primer name: Sequence 5' -> 3'	Reference	Taxon
R2IF1: AAGCARGGNGAYCCNCTNTC	Kojima and Fujiwara (2005)	*Baa; Bgg; Bgb; Bgm*
28S-F2: GTCAAAGTGAAGAAATTCAACGAAG	Mingazzini et al. (2011)	*Baa; Bgg; Bgb; Bgm*
28SFBGG: GAATCCGACTGTCTAATTAAAACAAAG	this study	*Baa; Bgg; Bgb; Bgm*
28S-R: TCCATTGCTGCGCGTCACTAATTAGATGAC	Bonandin et al. (2014)	*Baa; Bgg; Bgb; Bgm*
BggBdin3: TATGATATTCAACATGGTTATTG	Mingazzini (PhD thesis, 2011)	*Baa; Bgg*
BggArin: CGATCTAGTAGGTACTTCACACATGCA	this study	*Baa; Bgg*
Ins2_F: GGACAAGCGCACAGAGTCAG	this study	*Bgg; Bgb; Bgm*
Ins2_R: GGACATCCTGTTGGCGATTAC	this study	*Bgg; Bgb; Bgm*
BaadinATT ACCTCTCGCTGGCTCATTAC	this study	*Baa*
Bgg2rin: AAAAAGTCCTCTAACACTTGTATCTTC	this study	*Bgg*
Bgg2din: GTTAGTTTTACCCTACTGACGACCA	this study	*Bgg*
Bgg3rin: ACCCCAAAACAAGGAAGGAG	this study	*Bgg*
Bgg3din: TAGTATGGATTCGAGGATTAAGGAA	this study	*Bgg*
Bgg4rin: GGCGACCTATTACGATTTCAAGTAT	this study	*Bgg*
Bgg4din: GCACAAGTCTTTTTGGTTCATAGAT	this study	*Bgg*
Bgbrin: GGTCTCTCCATTCCCTCCTATC	this study	*Bgb*
Bgbdin: CCACTCCATTCAATACAGCATCT	this study	*Bgb*
Bgb2rin: CACCGTCTTTCTTTGGTATTAGTGT	this study	*Bgb*
Bgb2din: AGCTCGCTCTTCCTCCTCTC	this study	*Bgb*
Bgb3din GATCTTCACCTCCCTCATCATCAAT	this study	*Bgb*
Bgmrin: CTCCATTCAATACAGCATCTCC	this study	*Bgm*
Bgm2rin: CAGTGGGTTCGCTAGATAGTAGGT	this study	*Bgm*
Bgm2din: TTAAGGGGTCTAAGAAGGTGGAG	this study	*Bgm*
Bgm3rin: GTGTGTTTAATCGCCGTACAGTT	this study	*Bgm*
Ins_F: CTTTTGGAAGCCCTTGACCC	this study	*Bgb; Bgm*

Table 2. Nucleotide and amino acid (in parentheses) p-distances among the *Bacillus* R2 elements (below diagonal). Standard errors are shown above diagonal.

	1	2	3	4	5	6	7	8
1. R2_*Br*^{tunz}	-	0.005 (0.014)	0.004 (0.009)	0.004 (0.010)	0.005 (0.011)	0.007 (0.017)	0.011 (0.014)	0.008 (0.018)
2. R2_*Br*^{deg}	0.076 (0.164)	-	0.006 (0.015)	0.006 (0.015)	0.006 (0.015)	0.008 (0.017)	0.011 (0.015)	0.008 (0.018)
3. R2_*Bgm*	0.038 (0.077)	0.090 (0.181)	-	0.001 (0.003)	0.003 (0.007)	0.007 (0.016)	0.011 (0.014)	0.008 (0.017)
4. R2_*Bgm*^{del}	0.041 (0.085)	0.092 (0.186)	0.004 (0.008)	-	0.003 (0.007)	0.007 (0.016)	0.011 (0.014)	0.008 (0.017)
5. R2_*Bgb*	0.052 (0.101)	0.097 (0.189)	0.020 (0.035)	0.021 (0.036)	-	0.008 (0.016)	0.011 (0.014)	0.008 (0.017)
6. R2_*Bgg*^{A}	0.176 (0.271)	0.181 (0.304)	0.180 (0.282)	0.182 (0.284)	0.184 (0.289)	-	0.011 (0.015)	0.006 (0.013)
7. R2_*Bgg*^{B}	0.565 (0.841)	0.569 (0.838)	0.569 (0.845)	0.571 (0.849)	0.571 (0.849)	0.559 (0.828)	-	0.01 (0.014)
8. R2_*Baa*	0.184 (0.297)	0.192 (0.325)	0.192 (0.306)	0.194 (0.311)	0.195 (0.311)	0.080 (0.132)	0.565 (0.838)	-

Table S1. List of R2 sequences used for the phylogenetic analysis

Host species	Reference
Acyrthosiphon pisum (aphid)	Genome sequence
Bombus terrestris (bumblebee)	Genome sequence
Bombus impatiens (bumblebee)	Genome sequence
Camponotus floridanus (ant)	Genome sequence
Drosophila melanogaster (fruit fly)	Stage DE, Eickbush TH (2009)
Drosophila simulans (fruit fly)	Stage DE, Eickbush TH (2009)
Harpegnathos saltator (ant)	Genome sequence
Kalotermes flavicollis LI(termite)	Ghesini S et al. (2011)
Lepidururs arcticus (tadpole shrimp)	Luchetti A et al. (2012)
Lepidurus couesii (tadpole shrimp)	Luchetti A et al. (2012)
Lepidurus apus lubbockii (tadpole shrimp)	Luchetti A et al. (2012)
Linepithema humile (ant)	Genome sequence
Limulus polyphemus (horseshoe crab)	Burke WD et al. (1999)
Megachile rotundata (bee)	Genome sequence
Nasonia vitripennis R2A(jewel wasp)	Stage DE, Eickbush TH (2010)
Nasonia vitripennis R2B(jewel wasp)	Stage DE, Eickbush TH (2010)
Nasonia giraulti R2C(jewel wasp)	Stage DE, Eickbush TH (2010)
Pogonomyrmex barbatus (ant)	Burke WD et al. (1993)
Pogonomyrmex barbatus B (ant)	Burke WD et al. (1993)
Popillia japonica (japanese beetle)	Burke WD et al. (1993)
Porcellio scaber (rough woodlouse)	Burke WD et al. (1999)
Reticulitermes balkanensis (termite)	Ghesini S et al. (2011)
Reticuitermes grassei (termite)	Ghesini S et al. (2011)
Reticulitermes lucifugus (termite)	Ghesini S et al. (2011)
Reticulitermes urbis (termite)	Ghesini S et al. (2011)
Schistosoma mansoni A (bloodfluke)	Kojima KK, Fujiwara H (2005)
Schistosoma japonicum (bloodfluke)	Genome sequence
Schmidtea mediterranea (freshwater planarian)	Genome sequence
Solenopsis invicta (ant)	Kojima KK, Fujiwara H (2005)
Tenebrio molitor A (mealworm)	Burke WD et al. (1998)
Tenebrio molitor B (mealworm)	Burke WD et al. (1998)
Tetranychus urticae (red spider)	Genome sequence
Tribolium castaneum A (beetle)	Genome sequence
Tribolium castaneum B (beetle)	Genome sequence
Tribolium castaneum C (beetle)	Genome sequence
Triops longicaudatus (tadpole shrimp)	Kojima KK, Fujiwara H (2005)
Triops cancriformis (tadpole shrimp)	Mingazzini V et al. (2011)

Table S2. Reverse Transcriptase Amino Acid Sequences Compared in Figure 2.

Label in Figure 4	Non-LTR_1 (species)	Non-LTR_2 (species)	Amino Acid Divergence Per Site	Estimated Age of Divergence (MYA)
1	R2 (*D. melanogaster*)[a]	R2 (*D. simulans*)[a]	0,034	2,3
2	R2 (*D. yakuba*)[a]	R2 (*D. melanogaster*)[a]	0,094	6
3	Jockey (*D. funebris*)[a]	Jockey (*D. melanogaster*)[a]	0,249	39
4	R2 (*D. mercatum*)[a]	R2 (*D. melanogaster*)[a]	0,301	39
5	R2 (*F. auricularia*)[a]	R2 (*D. melanogaster*)[a]	0,53	250
6	Lian (*A. aegypti*)[a]	LOA (*D. silvestris*)[a]	0,555	250
7	SART1 (*B.mori*)[a]	RT1 (*A. gambiae*)[a]	0,619	250
8	R1 (*Aphonopelma sp.*)[a]	R1 (*D. melanogaster*)[a]	0,652	600
9	R2 (*P. scaber*)[a]	R2 (*D. melanogaster*)[a]	0,655	600
10	AgamR4_Ele1.1 (*A. gambiae*)[b]	R4 (*B. mori*)[a]	0,583	340
11	AaegR4_Ele1.1 (*A. aegypti*)	R4 (*B. mori*)[a]	0,569	340
12	AgamR4_Ele1.1 (*A. gambiae*)[b]	R4 (*A. lumbricoides*)[a]	0,709	600
13	AaegR4_Ele1.1 (*A. aegypti*)[b]	R4 (*A. lumbricoides*)[a]	0,664	600
14	Ag-L2-1 (*A. gambiae*)[b]	L2 (*A. aegypti*) [b]	0,38	145
15	Ag-Jockey-7 (*A. gambiae*)[b]	Jockey-7 (*A. aegypti*)[b]	0,422	145
16	Ag-R1-7 (*A. gambiae*))[b]	R1 (*A. aegypti*)[b]	0,433	145
17	Ag-Jockey-1 (*A. gambiae*)[b]	Jockey-1 (*A. aegypti*)[b]	0,437	145
18	Ag-R1-1 (*A. gambiae*)[b]	R1 (*A. aegypti*)[b]	0,44	145
19	Ag-Loner-1 (*A. gambiae*)[b]	Loner (*A. aegypti*)[b]	0,452	145
20	Ag-L1-5 (*A. gambiae*)[b]	L1 (*A. aegypti*) [b]	0,455	145
21	Ag-CR1-20 (*A. gambiae*)[b]	CR1 (*A. aegypti*)[b]	0,471	145
22	Ag-Jen-1 (*A. gambiae*)[b]	Jen (*A. aegypti*)[b]	0,473	145
23	Ag-I-2 (*A. gambiae*)[b]	I-2 (*A. aegypti*) [b]	0,525	145
24	Ag-Outcast-6 (*A. gambiae*)[b]	Outcast-6 (*A. aegypti*)[b]	0,586	145
25	Q (*Anopheles*)[c]	Worf (*Drosophila*)[c]	0,682	260
26	TI (*Anopheles*)[c]	Worf (*Drosophila*)[c]	0,663	260
27	SoriCR I A (*Scolitantides*)[c]	CR I A (*Maculinea*)[c]	0,023	10
28	BmCR I B (*Bombyx*)[c]	OcaCR I B (*Oberthueria*)[c]	0,046	10
29	MteQ (*M. teleius*)[c]	Q (*Anopheles*)[c]	0,6	350
30	Jockey (*D. melanogaster*)[c]	Jockey (*D. funebris*)[c]	0,247	40
31	AMY (*B. mori*)[c]	Helena (*D. mauritiana*)[c]	0,524	350
32	JuanA (*Aedes*)[c]	Jockey (*D. melanogaster*)[c]	0,605	250
33	NLRI Cth (*Chironomus*)[c]	Jockey (*D. melanogaster*)[c]	0,56	250
34	NLRI Cth (*Chironomus*)[c]	AMY (*B. mori*)[c]	0,674	350
a	R2 (*Bgb*)[d]	R2 (*Bgm*)[d]	0,011	1,9
b	R2 (*Bgb*)[d]	R2 (Bgm^{del})[d]	0,09	1,9
c	R2 (*Bgg*)[d]	R2 (*Baa*)[d]	0,109	15,37
d	R2 (*Bgm-Bgb*)[d]	R2 (*Bgg*)[d]	0,193	16,96
e	R2 (*Bgm-Bgb*)[d]	R2 (*Baa*)[d]	0,18	16,96
f	R2 (*Bgm*)[d]	R2 (*Br*)[d]	0,056	22,79
g	R2 (*Bgb*)[d]	R2 (*Br*)[d]	0,049	22,79
h	R2 (*Bgg*)[d]	R2 (*Br*)[d]	0,182	22,79
i	R2 (*Baa*)[d]	R2 (*Br*)[d]	0,171	22,79
l	R2(Bgm^{del})	R2(*Br*)	0,054	22,79

[a] Malik and Eickbush 1999
[b] Biedler et al. 2015
[c] Novikova et al. 2007
[d] Mantovani et al. 2001 and this work

2.2. Main results in *Triops cancriformis*.

With the aim to compare different organisms sharing the same reproductive strategies (even if at the species level) I have also analyzed the TEs content and composition in another invertebrate model represented by *T. cancriformis* (Crustacea Branchiopoda) sharing with *Bacillus* instances of bisexual and parthenogenetic reproduction.

The genomic library of *T. cancriformis* parthenogenetic Italian population, which I produced through the NGS technology for Ion Torrent sequencer, shows 10706 contigs containing TEs, corresponding to the 20% of the library, and 13924 contigs in which simple repeats of 1-7 bp are present, possibly representing microsatellites. Both TE classes are present with different percentages (class I, 11.4% and class II, 8.9%). For class I, a greater number of contigs shows the presence of Gipsy family elements, while for class II the Mariner family elements are the most represented.

These results are in contrast with the theories by which the accumulation and proliferation of TEs in unisexual organisms would lead to extinction of the host lineage or that, in the long run, TEs would be expected to be eliminated by genome purging, resulting in a genome free of such elements (Hickey, 1982).

The results are considered in detail in the following paragraph.

2.2.1. TEs content and composition in *Triops cancriformis*

Triops cancriformis is a species of tadpole shrimp found from Europe to the Middle East and India. As already defined in the paragraph 1.5.1., its populations can be gonochoric, parthenogenetic or hermaphroditic. I isolated genomic DNA using part of the cephalon/pleon, from samples of an Italian parthenogenetic population from Novara through phenol-chloroform protocol. I produced whole genome libraries through the NGS technology (Next Generation Sequencing) of Ion Torrent sequencer. Sequencing reads were generated by IonTorrent PGM using a 318v1 chip and a 400-bp sequencing kit. Sequencing produced 6,171,374 reads. For reads assemblies we collaborated with the bioinformatics group of Professor Rita Casadio (University of Bologna). The assembly was performed with MIRA 3.4.1.1 software (Chevreux et al., 1999). The number of reads assembled was 5172167, resulting in an average total coverage of 10.8x. Contigs obtained after the assembly were 52884; the N50 statistic was equal to 5306. TE identification was performed by RepeatMasker program (Smit et al., 1996). I found 10706 contigs (corresponding to the 20% of the library) containing TEs, while in 13924 contigs simple repeats of 1-7 bp (possibly representing microsatellites) occur. The most

represented TE class were retrotransposons (11,4%) while DNA transposons occur at 8,9%. The most frequent subclass among retrotransposons is given by LTR elements (9,9%), in contrast with the non-LTR retrotranposons that occurred only at 1,5%. Table 1 shows in detail the TEs families scored in the *Triops* library.

TE class/superfamily	TE family	N°
ClassI		
	LTR	2704
	Copia	124
	DIRS	2
	Gypsy	1569
	Pao	839
	Non-LTR	163
	CR1	60
	CRE	1
	Jockey	242
	LOA	58
	R1	90
	R2	14
	R4	6
	RTE	11
	L1	33
	L2	82
	SINE	17
ClassII		
	DNA	1806
	CMC	45
	Ginger	1
	hAT	1083
	Kolobok	3
	MULE	27
	P	103
	PIF	74
	PiggyBac	6
	Sola	23
	TcMar	1372
	Zator	7
	Maverick	83
	Helitron	58

Tab.1 Transposable element families found in the *Triops cancriformis* library. The contigs number in which TEs families are present (N°) is shown.

These data are very important because, at this time, studies addressed to confirm the theory that taxa with unisexual reproduction – in contrast with their close relative sexual species - could have an increase of TEs density, are hindered by the lack of suitable model organisms. In fact, the ideal model organism to identify consequences related directly to the reproduction mode and no to species-specific effects, should be a taxon in

which unisexual and sexual sister lineages share the same life-history, the same population size and the same selective pressure. In addition, there is the difficulty in sequencing and assembly of the TEs whole genomic load. For these reasons, studies addressing this topic are few and present opposing patterns: for example, in comparative analyses between selfing and outcrossing lineages in *Arabidopsis* and *Caenorhabditis* showed that TEs segregate at higher frequencies and persist longer in selfing populations. On the contrary, obligate parthenogenetic lineages of *Daphnia pulex* seem to have lower copy numbers of both TEs classes than cyclic parthenogenetic organisms of the same species (Dolgin et al. 2008; Lockton & Gaut 2010; Schaack et al. 2010).

Ricci et al. (2013) carried out a TEs genomic sequence survey by randomly cloning genomic fragments or/and Southern blot analyses of the three parental species *B. rossius*, *B. grandii*, and *B. atticus.* Although obtaining a low coverage sequencing, their data might represent a first general overview of the TEs situation in *Bacillus*. The parthenogenetic *B. atticus* shows the presence of 19 TE families; on the contrary in *B. rossius* and *B. grandii* a lower number of families were found. The absence of some families in the gonochoric genomes of *B. rossius* and *B. grandii* can be explained by the greater ability of bisexuals to oppose the TEs proliferation. Generally, DNA transposons were the most present in all libraries.

Bdelloidea rotifers are the only organisms whose ancestral condition is unisexuality. Early studies had demonstrated the total absence of retrotranposons in their genomes (Arkhipova and Meselson, 2000). This condition has been explained with the hypothesis that these organisms are more tolerant toward elements that move by excision of the previous copy than elements that produce copies of themselves, thereby increasing their content in the genome (Arkhipova and Meselson, 2000). Thereupon, other analyses showed the presence of the *Athena* non-LTR retrotransposons in telomeric/subtelomeric regions, but not in genes rich regions (Arkhipova et al., 2003; Gladyshev e Arkhipova, 2010). Kraaijeveld et al. (2012) have sequenced the genome of the parasitoid *Leptopilina clavipes* in bisexual reproducing populations and in parthenogenetic ones. The TEs copy number analysis in the two genomes showed that the copy number of DNA transposon was higher in parthenogenetic populations, while bisexual reproducing populations had more copies of non-LTR transposons. However, the most complete and recent study of Bast et al.(2015) has demonstrated that there is no link between reproductive mode and TEs whole-genome loads. They analysed five asexual arthropod lineages and their sexual

relatives: crustaceans (with two *Daphina pulex* lineages); insects (with one *Leptopilina clavipes* lineage) and chelicerates (with *Platynothrus peltifer*, *Hypochthonius rufulus*, *Steganacarus magnus*, and *Achipteria coleoptrata*). Their analyses showed no evidence for an increased TE load in genomes of asexual as compared to sexual lineages, neither for all TE classes nor for specific TE families. Suggesting that nonrecombining genomes do not accumulate TEs as predicted by theoretical expectations.

My data showed the presence of all TE classes in *T. cancriformis* library in which, however, retrotransposons are more represented. The higher presence of class I elements can be explained in many ways. One aspect is the induction of retrotransposons in response to stress. Hoffmann and Parsons (1997) defined the stress as "any environmental change that drastically reduces the fitness of an organism". Heat shock, gamma-irradiation, etc. are events that causing retrotransposons mobilization, induce their transposition. These events have been well documented in some species of *Drosophila* and in some plants (Wessler 1996; Vasil'eva et al. 1998). It is possible that a constitutive stress response over millions of years may have played a role in generating the high content of retrotransposons in *T. cancriformis.* Another line is the parasitic DNA hypothesis (Doolittle and Sapienza, 1980): the TEs copy number in a population depends on their ability to self-replicate. Thanks to their "copy-and-paste" transposition mechanism, retrotransposons are more able to increase in copy number than DNA transposons which move by "cut-and-paste" mechanism. However, the most probable evidence for the increase of retrotransposons is due to the parthenogenetic reproduction adopted by the *Triops cancriformis* population analyzed. The control mechanism of TEs proliferation is different in genomes with and without meiotic recombination. The independent assortment and recombination of chromosomes during meiosis can lead to the loss of TE copies and therefore to a decrease in their copy number, while unisexual reproduction would lead to an increase of these elements density due to the inability to eliminate them (Schaack et al., 2010).

2. ABROAD PERIOD

Next-generation sequencing (NGS) technologies allow to have a large quantity of nucleic acids sequences in a quicker and cheaper way than the previously used Sanger sequencing. This is leading to expansion at an unprecedented pace of genomic and transcriptomic science fields.

To learn how to analyze Next Generation Sequencing data, I spent my abroad period at the School of Biological, Biomedical and Environmental Sciences of the Hull University (U.K.), under the supervision of Professor Dave Lunt. My work there focused on the identification and comparison of some genes involved in meiotic processes in genomes of five species of the *Meloidogyne* genus, either reproducing through automictic parthenogenesis (*M. hapla, M. floridensis* and *M. chitwoodi*) or with an apomictic one (*M. incognita* and *M. jovanica*).

Some studies have shown that the "antiquity" of the loss of meiosis and then the occurrence of unisexual reproduction can be analyzed through the check of the simultaneous presence/absence of several meiotic genes and their “status” (Normark et al., 2003; Schurko and Logsdon, 2008). Often when a cell function is lost or becomes non-essential, the selective pressures acting on the genes required for this function decrease and these genes will thus be free to accumulate deleterious mutations that can lead to a very degraded sequence or even to loss of function variants. Therefore, the sequence characterization of meiotic genes in addition to the study of the evolutionary rates and the functions of these genes, also allow to identify the loss of meiosis in populations.

I first found, from the literature, ten meiotic genes that could have been good candidates for the analysis. The dataset was split into highly conserved meiotic genes, present in all eukaryotes, and specific meiotic genes of the phylum Nematoda. With the Python programming language I learned to handle the vast amount of data from databases such as GenBank and WormBase, from which I downloaded all the amino acid sequences of interest. I used the method based on homology searching through PSI-BLAST to detect the presence of each gene within the *Meloidogyne* genome. The query model, for this research, was created using the position specific scoring matrix (PSSM) derived from the alignment of the protein sequences for each gene.

The first results showed the presence of all the genes considered in all genomes, without any apparent distinction between automictic and apomictic organisms. This could be due to the likely recent origin of the apomictic species and their recent speciation (2-4 million

years ago). Further analyses are ongoing to determine the evolutionary model, the phylogenetic relationships and selective pressures agents for each group of genes.

The new knowledge that I acquired during my time in the laboratory of Professor D. Lunt enabled me to analyze Next Generation Sequencing data from *T. cancriformis* genome (see paragraph 3.1.) but also will enable me to investigate on the evolution of meiotic genes in the species of *Triops* and *Bacillus* genus that are of interest.

3. CONCLUSIONS

TEs are an important fraction of many living organisms genomes.

During these years many studies aimed to understand the precise mechanisms by which TE activity and copy number are regulated were made, but relatively few about the dynamics between TEs and their host genome. Hickey in 1982 suggested that the reproductive strategy of the host can influence the TEs evolutionary dynamics.

Genetic variability, typical of bisexually reproducing organisms, is the result of mutations, crossing-over, segregation, independent assortment and fertilization. All these phenomena can lead to dispersal and fixation of new TE variants within a population, but for the same phenomena a new TE variant can be eliminated. In this frame, TEs create genetic changes, such as heterozygous insertions, upon which natural selection has more opportunity to operate. In contrast, unisexual organisms should not effectively contrast the TEs proliferation due to the inability to eliminate them through exclusive mechanisms of sexual reproduction. In addition, in these organisms the new TE insertions will also tend to become homozygous and natural selection may not be so effective as in bisexual populations. In these organisms TEs would be accumulated even up to the host lineage extinction (Hickey, 1982; Schaack et al., 2010).

Under these assumptions, the common expectation should be that the genomes of unisexual organisms are TEs-free, or at least lack of functional elements, to avoid their excessive accumulation and consequently the extinction of the organism, as explained by the Muller's ratchet theory (Muller, 1964; Arkhipova and Meselson, 2000).

The data of my work are in contrast with these expectations because they show a significant presence of functional R2 retrotransposons in both facultative and obligatory parthenogenetic genomes of *B. rossius* and *B. atticus* and also a significant presence of TEs in the parthenogenetic *T. cancriformis* library.

In the *Bacillus* genus I focused on the R2 non-LTR retrotransposon. It is one of the most site-specific retrotransposon family studied and has wide distribution in metazoan genomes. Target site selection appears as a strategy to optimize the element-host relationship. In particular, the insertion into tandem units repeats appears more efficient because when selective pressure eliminates inactive rDNA units, often by unequal crossover between sister chromatids, new rDNA units are replaced through compensatory genomic turnover mechanisms, leading to new sites for further R2 insertions (Eickbush e Eickbush, 2007). The relationship between R2 and its host is quite interesting because

allows to study also the dynamics between TE and its genomic niche.

I isolated and sequenced by primer walking seven R2 complete elements: one in *B. rossius redtebacheri* (gonochoric population); one in *B. atticus atticus* (obbligatory parthenogenetic population); two in *B. grandii grandii* (gonochoric population); one in *B. grandii benazzii* (gonochoric population) and two in *B. grandii maretimi* (gonochoric population).

The R2 element of *B. rossius redtebacheri* (R2Br^{deg}) appear degenerate, its ORF showing 14 frameshift mutations and one stop codon. This R2 degenerate variant is not private of parthenogenetic populations; in fact, the R2 population survey on functional (R2Br^{fun}) and degenerate (R2Br^{deg}) variants showed that both variants are present in all *B. rossius* populations (either parthenogenetic or bisexual) with different percentages. In addition, the presence of R2Br^{deg} in *B. rossius tripolitanus* A indicates that this degenerate element was already present in the ancestral *B. rossius* genome. When the separation of Europe and North Africa led to allopatric *B. rossius* lineage break (>5 Myr ago; Mantovani et al., 2001). Sequence variability analyses between the functional and degenerate dataset and between parthenogenetic and gonochoric populations have indicated that no link exists between sequence variability and reproductive strategies in *Bacillus.* The peculiar distribution of the two R2 variants, present in all *B. rossius* subspecies but with different percentages, can be explained by the coexistence of a collection of R2 lineages inherited from a common ancestral genome and by the selection/expansion of some lineages respect the others. This scenario is suggested by the “library hypothesis” that predicts the presence of a library of low-copy satellites (but applicable to tandem repeated families in general) within related taxa and their different amplification in each species lead to have one or a few of them as major satellites present in the genome (Salser et al. 1976). This hypothesis has been first demonstrated for four satellites in the congeneric species *Palorus* (Mestrovic et al., 1998; Mravinac et al., 2002) and for the Bag320 satellite in *Bacillus* stick insects (Cesari et al., 2003). As suggested by Luchetti and Mantovani (2013), even R2 elements distribution could be explained by this hypothesis because they are inserted in tandem-repeats array, the rDNA, and may be subject to mechanisms ruling the evolution of their niche (Zhang et al., 2008). For these reasons it can be hypothesized the existence of a library of R2 elements inserted-specific into 28S ribosomal genes within an ancestral genome and their different amplification in each species.

Often, some TEs are present in the host genomes with a mixture of active and degenerated copies (Lerat et al., 2003). There are several hypotheses that try to explain

the persistence of degenerate elements in the genome. A probable hypothesis in my case could be the target site selection of R2 elements. The R2 site specific insertion strategy could allow degenerate variant to persist in the *Bacillus* genome because they could duplicate via genomic amplification of inserted 28S by unequal recombination and the molecular drive's homogenizing forces could allows to maintain the silenced or degenerate elements over evolutionary time.

The another R2 elements which I isolated from *B. atticus* and *B. grandii* subspecies show the typical structure of the element, with a N-terminal domain that contain one ZF motif (except for R2Bgg^{B}), a central RT domain and a C-terminal domain. However, the R2Bgg^{B} sequence is very divergent from the others, the presence of two ZF motifs and a nucleotide divergence greater than 50% could *i*ndicate a different lineage of R2 element.

The coexistence of multiple lineages of R2, as well as the extinction of some ones, can be explained by the genomic control mechanisms, population dynamics of the host species or by their transposition mechanism. Previous studies showed that the R2 elements with different ZF motifs have a different DNA binding mode , suggesting that a certain plasticity in the integration mechanism exists (Thompson e Christensen, 2011).

In addition, my analyses for the first time suggest that also R2 retrotransposons can use HT as a strategy to colonize a new genome. HT is considered an essential part of the life cycle of TEs as it allows them to colonize new genomes and to escape vertical extinction resulting from elimination, by drift or selection, or inactivation due to mutational decline (Schaack et al., 2010). Many cases of HT have been demonstrated for many TEs but extensive previous studies on R2 element in the *Drosophila* genomes have always shown a vertical transmission of this element since the Radiata-Bilateria split (Eickbush & Eickbush, 1995; Kojima and Fujiwara 2005). However, the observed lower divergence among *B. rossius, B. g. maretimi* and *B. g. benazzii* R2 elements with respect to the divergence time of the hosts, the phylogenetic incongruence between R2 and its host and the patchy distribution of the R2 within *Bacillus* genus are evidences of HT instance.

It is very difficult to put forward hypotheses on HT events especially because the mechanisms by which TEs can be transported between organisms is still unknown. Some mechanisms hypothesized involve different possible vectors or some intrinsic characteristics of certain LTR retrotransposons; for example the Gipsy elements have the characteristic to encode a protein similar for structure and function to the envelope protein present in retroviruses. This protein allows them to move easily between species by infecting new cells (Havecker et al., 2004). However, the most simple case of possible HT

event occur when two different species hybridize; in fact, the hybridization lead to introgression of genetic material. Often, especially if reproductive isolation does not occur, the hybridization between two taxa is followed by backcrossing between the hybrids and parental lineages. After several generation of backcrossing, parental-like individuals can be found in which a small proportion of the alien genome is present. This latter is the most plausible case to explain R2 HT in *Bacillus* species; in fact, *B. rossius* and *B. g. benazzi* are the parental species of the hybridogenetic strain *B. rossius* - *grandi benazzi* (Scali et al., 2003). The hybridogenetic strain share the same geographic area with its parental species (see Fig. 11) this making gene flow possible. In addition, in these hybridogenetic lineages of *Bacillus* it was observed, for the first time in the animal kingdom, the androgenesis phenomenon in which the embryo derives from the paternal nuclear genome only, having eliminated the maternal pronucleus. The formation of a diploid nucleus takes place by fusion of two male pronuclei. This could re-forming the *B. grandii benazzii* paternal species.

The studies on the relationship between TEs activity and host reproductive strategies are very few because it is not simple found the suitable model organisms and it is difficult to sequence and assembling the TEs from genome-wide data. In fact, most analyses of TEs comparison in populations with different reproductive strategies were made taking into consideration only one class or few families of TEs. For example the study of Dolgin et al. (2008) on *Caenorhabditis* genus took in consideration only Tc1 DNA transposon family; Lockton & Gaut (2010) worked on three class I retrotransposons (LINE, SINE and *Gypsy*) and four class II DNA transposons (*Ac, Basho*, MITE and CACTA) in *Arabidopsis* genus; Schaack et al. (2010) analyzed six families of DNA transposons previously identified in the *D. pulex* genome. Only Kraaijeveld et al. (2012) had made a genome-wide TE copy number and then they focused on gipsy-like elements. All these analyses give us information on how the different reproductive modes may influence the activity of those particular TEs families. Different TE types have different mechanisms through which they spread within a genome, affecting their population dynamics. Therefore, the approach most complete and thorough to investigate this relationship is the analysis of TEs whole genomic load. Bast et al. (2015) in fact used this approach to investigate the TEs loads in five independently derived asexual arthropod lineages and their sexual relatives, thus to be considered the most complete and exhaustive work.

My data showed a TEs overview in a *T. cancriformis* parthenogenetic population. In contrast with the theoretical expectation that unisexual genomes should be TEs-free, or at

least lack of functional elements, to avoid their excessive accumulation and consequently the extinction of the organism, my analyses highlighted that the 20% of the library is composed by TEs, in which both TEs classes are widely represented (class I, 11.4% and class II, 8.9%). For the class I, a greater number of contigs shows the presence of Gipsy family elements, while for class II the Mariner family elements are the most represented. The most frequent subclass among retrotransposons is given by LTR elements (9,9%), in contrast with the non-LTR retrotransposons that occurred only at 1,5%.

However, Repeat Masker program, that I used for TEs identification, has been shown to be very fast and rather efficient, yet it is based on homology search, therefore it cannot detect completely novel elements but only are already known to exist. For this reason, the TEs percentage may be underestimated. However, different studies confirm that this program is helpful for a first step in the identification of repeats (Lerat, 2009).

The identification of repeats in genomic sequences is very difficult. TEs are present in the genome for a long time, and often copies of the same family can be highly modified due to point mutations, rearrangements or deletions in their sequence, consequently the elements are difficult to identify with only similarity approaches. Also it is difficult to determinate the real boundaries of the sequences, usually some TEs insert into other transposon sequence, forming nested elements (Kaminker et al., 2002). These considerations lead to prefer multiple approaches for the TEs identification. As I said before, a first approach can be done using a program based on sequence homology search which through a similarity search based on local alignments identifies TEs belonging to known classes. After this, there are many different programs that can be used to find new elements: programs to find non-LTR retrotransposons, which search the presence of a polyA tail in 3′ of the sequence and of target site duplications; programs to find LTR retrotransposons, which search the presence, size range and the percentage identity between the two long terminal repeat (LTR); and other several programs to find the major TEs families. For *de novo* identification of TEs an approach is to take advantage of the repetitive nature of TEs, without considering particular motifs or structures. There are different programs to detecting *de novo* TEs, some of these detect TEs using a sequence similarity scoring system which uses BLAST (Altschul et al., 1990) to perform the self-comparison and then uses clustering method to form repeat families. Instead other programs searching the occurrence of small repeated motifs.

My future perspectives are to use these different programs in the *T. cancriformis* library in order to estimate a TEs genomic composition as realistic as possible and also to

investigate the relationship between TEs dynamics and host reproductive strategies through genome-wide sequencing of gonochoric and hermaphroditic *T. cancriformis* populations in order to able to compare the TEs presence in genomes with different reproductive strategies.

4. REFERENCES

Alberola T. M., Bori L. and de Frutos R. 1997. Structural analysis of *Drosophila* subobscura gypsy elements (gypsyDs). Genetica. 100(1-3):39-48.

Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. 1990. Basic local alignment search tool. J Mol Biol 215: 403–410.

Arkhipova I. R. and Meselson M. 2000. Transposable elements in sexual and ancient asexual taxa. Proc. Natl. Acad. Sci. USA 97: 14473-77.

Arkhipova I. R., Piatkov K.I., Meselson M. and Evgen'ev M. B. 2003. Retroelements contain intron in diverse invertebrate taxa. Nature genetics. Vol. 33

Arkhipova I. and Meselson M. 2005 Deleterious transposable elements and the extinction of asexuals. BioEssays, 27, 76–85.

Biedler J. K., Chen X. and Tu Z. 2015 Horizontal transmission of an R4 clade non-long terminal repeat retrotransposon between the divergent *Aedes* and *Anopheles* mosquito genera. Insect Molecular Biology 24(3), 331–337.

Boeke, J. D., Garfinkel D. J., Styles C. A., and Fink G. R. 1985 Ty Elements Transpose through an RNA Intermediate. Cell 40: 491–500.

Bonandin L., Scavariello C., Luchetti A. and Mantovani B. 2014 Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta, Phasmida). Insect Molecular Biology 23(6): 808-820.

Bonandin L. 2016 Sex and repetitive sequence dynamics in *Bacillus* stick insects (Phasmida, Bacillidae). PhD thesis.

Brock, P. D. 2011. Order Phasmida Leach.1815. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. File online: http://Phasmida.SpeciesFile.org

Bunikis J. and Barbour A. G. 2005 Ticks have R2 retrotransposons but not the

consensus transposon target site of other arthropods. Insect Mol Biol 14: 465–474.

Burke W.D., Carolyn C.C. and Eickbush T.H. 1987 The Site-Specific Ribosomal Insertion Element Type II of *Bombyx mori* (R2Bm) Contains the Coding Sequence for a Reverse Transcriptase-Like Enzyme. Mol Cell Biol. 7(6):2221-30.

Burke W. D., Eickbush D.G., Xiong Y., Jakubczak J. L., and Eickbush, T. H. 1993 Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. Mol. Biol. Evol. 10,163–185

Burke W.D., Malik H.S., Lathe W.C. III, Eickbush T.H. 1998 Are retrotransposons long-term hitchhikers? Nature 392(6672):141-2.

Burke WD, Malik HS, Jones JP, Eickbush TH. 1999 The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol. Biol. Evol. 16(4):502-11.

Burwinkel B. and Kilimann M. W. 1998 Unequal homologous recombination between LINE1 elements as a mutational mechanism in human genetic disease. Journal of Molecular Biology 277(3):513-7.

Capy P., R. Vitalis, T. Langin, D. Higuet and C. Bazin 1996. Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? J. Mol. Evol. 42:359–369.

Cesari M., Luchetti A., Passamonti M., Scali V. and Mantovani B. 2003. PCR amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). Gene 312:289–295.

Cesari M., Mularoni L., Scanabissi F. and Mantovani B. 2004. Characterization of dinucleotide microsatellite loci in the living fossil tadpole shrimp *Triops cancriformis* (Crustacea Branchiopoda Notostraca). Molecular Ecology Notes, 4, 733–735.

Charlesworth D. and Wright S. I. 2001. Breeding systems and genome evolution. Curr Opin Genet Dev 11: 685–690

Chénais B., Caruso A., Hiard S. and Casse N. 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to

stressful environments. Gene. 509(1):7-15.

Chevreux B., Wetter T. and Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. Comput. Sci. Biol.: Proc. German Conference on Bioinformatics GCB'99 GCB, pp. 45–56.

Christensen S.M., Ye J. and Eickbush T.H. 2006. RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc. Natl. Acad. Sci. U.S.A. 103: 17602–17607.

Cohen S. M.,Cappello J., Lodish H. F., 1984. Transcription of *Dictyostelium discoideum* transposable element DIRS-1.Mol Cell Biol. 4(11): 2332–2340.

Cordaux R. and Batzer M. A. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 10(10): 691-703.

Crick F., 1979 Split genes and RNA splicing. Science; 204(4390):264-71.

Daniels S. B. , Peterson K. R., Strausbaugh L. D., Kidwell M. G. and Chovnick A. 1990 Evidence for the horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics. 124:339–355.

Dawkins R., 1976 The selfish gene. Published by Oxford University Press

DeBerardinis R. J. and Kazazian H. H. Jr. 1999. Analysis of the promoter from an expanding mouse retrotransposon subfamily. Genomics 56: 317–323.

Dolgin E. S. and Charlesworth B. 2006 The fate of transposable elements in asexual populations. Genetics, 174, 817–827.

Dolgin E. S., Charlesworth B., Cutter A. D. 2008 Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis* nematodes. Genetics Research, 90, 317–329.

Doolittle, W. F., and C. Sapienza, 1980 Selfish genes, the phenotype paradigm and genome evolution. Nature 284: 601–603.

Eickbush D. G. and Eickbush T. H. 1995 Vertical transmission of the retrotransposable

elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. Genetics 139: 671–684.

Eickbush T. H. and Eickbush D., **2007**. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175: 477- 85.

Eickbush T. H. and Jamburuthugoda V. K. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134: 221–234.

Eickbush D.G. and Eickbush T.H., 2010. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. Molecular and cellular biology 30(13): 3142-3150.

Eickbush D. H., Burke W. D. and Eickbush T.H. 2013. Evolution of the R2 retrotransposon ribozyme and its self-cleavage site. PloS One 16;8(9):e66441.

Fedor M. J. 2009. Comparative enzymology and structural biology of RNA self-cleavage. Annu Rev Biophys 38: 271-299

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.

Ferré-D'Amaré A. R., Zhou K., Doudna J. A. 1998. Crystal structure of a hepatitis delta virus ribozyme. Nature 395: 567-574.

Flavell A. J., 1981 Did retroviruses evolve from transposable elements? Nature 289:10–11.

Ganley A. R. and Kobayashi T. 2007 Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res. 17(2):184-91.

Ghesini S., Luchetti A., Marini M. and Mantovani B. 2011 The Non-LTR retrotransposon R2 in termites (Insecta, Isoptera): characterization and dynamics. J Mol Evol 72: 296–305.

Ghiselli F., Milani L., Scali V., Passamonti M. 2007 The *Leptynia hispanica* species complex (Insecta Phasmida): polyploidy, parthenogenesis, hybridization and more. Molecular Ecology 16: 4256-4268.

Gladyshev E.A. and Arkhipova, I.R., 2010. Subtelomeric non-LTR retrotransposon Hebe in the bdelloid rotifer *Adineta vaga* is subject to inactivation by deletions but not 5' truncations. Mobile DNA 1: 12.

Gould S. J. and Vrba E. S. 1982. Exaptation - a missing term in the science of form. Paleobiology 8 (1): 4–15.

Grebelnyi S.D. 2009 Unisexual reproduction: either maternal or paternal inheritance. Russian Journal of Developmental Biology, 39 (3): 145-149.

Havecker E. R., Gao X., Voytas D. F. 2004 The diversity of LTR retrotransposons. Genome Biol 5: 225

Hickey Donal A., 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 101(3-4): 519–531.

Hoffmann A. A. and Parsons P. A. 1997 Extreme Environmental Change and Evolution. Cambridge University Press.

Horn R. L., Kuehn R., Drechsel V. and Cowley D. E. 2014 Discriminating between the Effects of Founding Events and Reproductive Mode on the Genetic Structure of *Triops* Populations (Branchiopoda: Notostraca). PloS One 13;9(5):e97473.

Ivancevic A. M., Walsh A. M., Kortschak R. D., Adelson D. L. 2013 Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. Bioessays. 35(12):1071-82. doi: 10.1002/bies.201300072

Jordan I. K., Rogozin I. B., Glazko G. V. and Koonin E. V. 2003 Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 19, 68–72.

Jurka J. 1997 Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA; 94:1872-1877.

Kaminker J. S., Bergman C. M., Kronmiller B., Carlson J., Svirskas R., Patel S. *et al*. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol 3: RESEARCH0084.

Kapitonov, V.V. and Jurka, J., 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 3, e181.

Kapitonov V.V. and Jurka J. 2006 Self-synthesizing DNA transposons in eukaryotes. Proc. Natl. Acad. Sci. U. S. A. 103, 4540–4545

Kapitonov V.V. and Jurka J. 2007 Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. 23(10):521-9

Kidwell, M. G., 1977 Reciprocal Differences in Female Recombination Associated with Hybrid Dysgenesis in Drosophila Melanogaster. Genetical Research 30:77–88.

Kidwell, M. G., 1979 Hybrid dysgenesis in Drosophila melanogaster: the relationship between the P-M and I-R interaction systems. Genet. Res. 33: 205–217.

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.

Kojima K. K., Fujiwara H. 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol 20:351–61.

Kojima K. K., Fujiwara H. 2005 Long-term inheritance of the 28S rDNA-specific retrotransposon R2. Mol Biol Evol 22:2157–65.

Kojima K. K., Kuma K., Toh H. and Fujiwara H., 2006 Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. Mol. Biol. Evol. 23: 1984–1993.

Korn M., Marrone F., Pérez-Bote J. L., Machado M., Cristo, M., Cancela da Fonseca L., Hundsdoerfer A. K., 2006. Sister species within the *Triops cancriformis* lineage (Crustacea, Notostraca). Zool. Scr. 35, 301–322.

Korn M., Green A. J., Machado M., Garcia-de-Lomas J., Cristo M., et al. 2010 Phylogeny, molecular ecology and taxonomy of southern Iberian lineages of *Triops*

mauritanicus (Crustacea: Notostraca). Org Divers Evol 10: 409–440.

Korn M., Rabet N., Ghate H. V., Marrone F. and Hundsdoerfer A. K. 2013 Molecular phylogeny of the Notostraca. Molecular Phylogenetics and Evolution 69 1159–1171

Kraaijeveld K., Zwanenburg B., Hubert B. 2012 Transposon proliferation in an asexual parasitoid. Molecular Ecology, 21,3898–3906.

Levin, H.L. and Moran J.V. 2011 Dynamic interactions between transposable elements and their hosts. Nat Rev Genet. 12(9): p. 615- 27.

Lerat E., Rizzon C. and Biémont C. 2003 Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. Genome Res. 13(8):1889-96.

Lerat E. 2009 Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity 104, 520–533.

Linder F. 1952. Contributions to the morphology and taxonomy of the Branchiopoda Notostraca, with special reference to the North American species. Proceedings of the United States National Museum 102: 1–69.

Lockton S. and Gaut B. S. 2010 The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. BMC Evolutionary Biology, 10, 10.

Longhurst A. R. 1955. A review of the Notostraca. Bull. Br. Mus. Nat. Hist. Zool. 3, 1–57.

Loreto E. L. S., Carareto C. M. and Capy P. 2008 Revisiting horizontal transfer of transposable elements in Drosophila. Heredity. 100:545–554.

Luan D.D. and Eickbush T. H. 1995 RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol 15:3882-3891.

Luchetti A., Cesari M., Carrara G., Cavicchi S., Passamonti M., Scali V. 2003. Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta, Phasmatodea). J. Mol. Evol. 56: 587–596.

Luchetti A, Mingazzini V. and Mantovani B. 2012 28S junctions and chimeric elements

of the rDNA targeting non-LTR retrotransposon R2 in crustacean living fossils (Branchiopoda, Notostraca). Genomics 100: 51–56.

Luchetti A., Mantovani B. 2013 Non-LTR R2 element evolutionary patterns: phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages. Plos One. 8(2):e57076. Doi: 10.1371

Lynch J. E. 1966. *Lepidurus lemmoni* Holmes: a redescription with notes on variation and distribution. Trans. Am. Microsc. Soc. 85, 181–192.

Malik H. S., Henikoff S. and Eickbush T. H. 2000 Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10:1307–1318.

Mantovani B., Scali V., 1990. Preliminary report on a hybridogenetic stick insect (Phasmatodea): the first case among invertebrates. Inv. Repr. Dev., 18: 185-188.

Mantovani B., Scali V., Tinti F. 1992 New morfphological and allozymic characterization of *Bacillus whitei* and *B. lynceorum* hybrid Complexs (Insecta Phasmatodea). Biol. Zent. 111: 75-91.

Mantovani B. and Scali V., 1992. Hybridogenesis and androgenesis in the stick insect *Bacillus rossius–grandii benazzii* (Insecta, Phasmatodea). Evolution 46: 783–96.

Mantovani B., Scali V., 1993. Genetic characterization of *Bacillus grandii maretimi* (Insecta Phasmatodea) in relation to its ecology. Vie Milieu. 43 (4): 241-246.

MantovaniB., Tinti F., Scali V., 1995. Genetic structure and similarity relationship of the unisexual Cyprian *Bacillus* (Insecta Phasmatodea). Biol. Zent.bl., 114: 299-306.

Mantovani B., Passamonti M. and Scali V. 1999. Genomic evolution in parental and hybrid taxa of the genus *Bacillus* (Insecta, Phasmatodea). It. J. Zool. 66: 265– 72.

Mantovani B, Passamonti M. and Scali V. 2001. The mitochondrial cytocrome oxidase II gene in *Bacillus* stick insect: ancestry of hybrids, androgenesis, and phylogenetic relationships. Mol. Phylog. Evol. 19 (1): 157-63.

Mantovani B., Cesari M., Scanabissi F., 2004. Molecular taxonomy and phylogeny of the ‘living fossil’ lineages *Triops* and *Lepidurus* (Branchiopoda: Notostraca). Zool. Scr. 33,

367–374.

Mantovani B., Cesari M., Luchetti A., Scanabissi F. 2008 Mitochondrial and nuclear DNA variability in the living fossil *Triops cancriformis* (Basc, 1801) (Crustacea, Branchiododa, Notostraca). Heredity 100: 496–505.

Mantovani B., Cesari M., Scanabissi F. 2009 Molecular taxonomy and phylogeny of Italian *Lepidurus* taxa (Branchiopoda: Notostraca). Italian J Zool, 76:358–365.

Marescalchi O., Cesari M., Eder E., Scanabissi F. and Mantovani B. 2005 Chromosomes in sexual populations of Notostracan and Con-chostracan taxa (Crustacea, Branchiopoda). Caryologia: International Journal of Cytology, Cytosystematics and Cytogenetics. Volume 58, Iusse 2, pag. 164-170.

Marino-Ramirez L, Lewis KC, Landsman D, Jordan IK. 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. Cytogenet Genome Res;110:333–41

Mathers T. C., Hammond R. L., Jenner R. A., Hänfling B. and Gómez A. 2013a. Multiple global radiations in tadpole shrimps challenge the concept of 'living fossils'. PeerJ 1, e62.

Mathers T. C., Hammond R. L., Jenner R. A., Zierold T., Hänfling B. and Gómez A. 2013b. High lability of sexual system over 250 million years of evolution in morphologically conservative tadpole shrimps. BMC Evolutionary Biology, 13:30.

Mathers T. C., Hammond R. L., Jenner R. A., Hänfling B., Atkins J. and Gómez A. 2015. Transition in sexual system and sex chromosome evolution in the tadpole shrimp *Triops cancriformis.* Heredity 115(1): 37–46.

McClintock B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36:344–355.

McClintock B. 1953. Induction of instability at selected loci in maize. Genetics 38 (6): 579–99

McClure M. A. 1991. Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol. Biol. Evol. 8: 835–856.

McDonald J. F. 1983. The molecular basis of adaptation: a critical review of relevant ideas and observations. Annu. Rev. Ecol. Syst. 14**:** 77–102.

McLean C., Bucheton A. and Finnegan D. J. 1993. The 5' untranslated region of the I factor, a long interspersed element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression. Mol. Cell. Biol. 13: 1042–1050.

Mestrovic N., Plohl M., Mravinac B. and Ugarkovic D. 1998. Evolution of satellite DNAs from the genus *Palorus* – Experimental evidence for the library hypothesis. Mol. Biol. Evol. 15: 1062–1068.

Milani L., Ghiselli F., Pellecchia M., Scali V. and Passamonti M. 2010. Reticulate evolution in stick insects: the case of *Clonopsis* (Insecta Phasmida). BMC Evolutionary Biology 10:258.

Mingazzini V., Luchetti A. and Mantovani B. 2011 R2 dynamics in Triops cancriformis (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution. Heredity. 106(4):567-75.

Mingazzini V., 2011. Transposable elements: structure and dynamic of the autonomous retrotransposon R2 in Arthropoda with non-canonical reproduction. P.hD. Thesis, University of Bologna.

Mravinac B., Plohl M., Mestrovic N. and Ugarkovic D. 2002. Sequence of PRAT satellite DNA ‘‘frozen” in some Coleopteran species. J. Mol. Evol. 54: 774–783.

Muller, H. J. 1964 The relation of recombination to mutational advance. Mut. Res. 1, 2 –9. (doi:10.1016/0027-5107(64)90047-8)

Nei M. and Rooney A. P., 2005. Concerted and birth-and-death evolution of multigene families. Ann Rev Genet 39: 121-52

Normark B. B., Judson O. P. and Moran N. A. 2003 Genomic signatures of ancient asexual lineages. Biological Journal of the Linnean Society, 79, 69–84.

Novikova O., Śliwińska E., Fet V., Settele J., Blinov A. and Woyciechowski M. 2007 CR1 clade of non-LTR retrotransposons from *Maculinea* butterflies (Lepidoptera:

Lycaenidae): evidence for recent horizontal transmission. BMC Evol. Biol. 7:93.

Peters J.E. 2014 Tn7. Microbiol. Spectr. 2:1–20.

Perez-Gonzalez C. E. and Eickbush T. H. 2001 Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans.* Genetics 158**:** 1557–1567.

Picard, G., J. C. Bre´gliano, A. Bucheton, J. M. Lavige, A. Pe´lisson et al., 1978 Non-Mendelian female sterility and hybrid dysgenesis in Drosophila melanogaster. Genet. Res. 32: 275–287.

Piégu B., Bire S., Arensburger P. and Bigot Y. 2015 A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. Mol Phylogenet Evol. 86:90-109.

Piskurek O. and Okada N. 2007 Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals.Proc Natl Acad Sci USA. 104(29):12046-51.

Pritham E.J. 2009 Transposable Elements and Factors Influencing their Success in Eukaryotes. J Hered. 100(5): 648–655.

Ricci M., Luchetti A., Bonandin L. and Mantovani B. 2013 Random DNA libraries from three species of the stick insect genus *Bacillus* (Insecta: Phasmida): repetitive DNA characterization and first observation of polyneopteran MITEs. Genome; 56(12):729-35.

Roiha H. and Glover D. M. 1981 Duplicated rDNA sequences of variable lengths flanking the short type I insertions in the rDNA of Drosophila melanogaster. Nucleic Acids Res. 9:5521-5532

Ruminski D. J., Webb C.H.T., Riccitelli N.J., Luptak A. 2011. Processing and translation initiation of non-long terminal repeat retrotransposons by Hepatitis Delta Virus (HDV)-like self-cleaving ribozymes. J. Biol. Chem. 286 41286–41295.

Salathé M., Kouyos R. D. and Bonhoeffer S. 2008. The state of affairs in the kingdom of the Red Queen. Trends Ecol. Evol. 23 (8): 439-45.

Salser W., Bowen S., Browne D. et al. (11 co-authors). 1976. Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed. Proc. 35:23–

35.

Sassaman C. 1991. Sex-ratio variation in female-biased populations of notostracans.

Hydrobiologia 212: 169–179.

Sassaman C., Weeks S. C. 1993. The genetic mechanism of sex determination in the Chonchostracan shrimp *Eulimnadia texana*. Am Naturalist 141: 314–328.

Sassaman C., Simovich M. A., Fugate M. 1997. Reproductive isolation and genetic differentiation in North American species of *Triops* (Crustacea: Branchiopoda: Notostraca). Hydrobiologia 359, 125–147.

Scali V., Tinti F., Mantovani B. and Marescalchi O., 1995. Mate recognition and gamete cytology features allow hybrid species production and evolution in *Bacillus* stick insects. In: Balletto E, Minelli A, eds. Speciazione: tempo e modi. Bollettino di Zoologia 62: 59–70.

Scali V., Passamonti M., Marescalchi O., Mantovani B., 2003. Linkage between sexual and asexual lineages: genome evolution in *Bacillus* sticks insects. Biol. J. Linn. Soc. 79: 137-50.

Scanabissi F., Eder E. and Cesari M. 2005. Male occurrence in Austrian populations of Triops cancriformis (Branchiopoda, Notostraca) and ultrastructural observations of the male gonad. Invertebr. Biol. 124: 57–65.

Schaack S., Choi E., Lynch M. and Pritham EJ. 2010. DNA transposons and the role of recombination in mutation accumulation in Daphnia pulex. Genome Biol.11:R46.

Schaack S., Gilbert C., Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25: 537–46.

Schurko A. M. and Logsdon J. M. Jr 2008. Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. Bioessays 30(6):579-89.

Sela N., Mersch B., Hotz-Wagenblatt A. and Ast G. 2010. Characteristics of transposable element exonization within human and mouse. PLoS One 5(6): e10907.

Sellick J.T.C., 1998. The micropylar plate of the eggs of Phasmida, with a survey of the range of plate form within the order. Systematic Entomology 23, 203–228.

Shapiro, J. A., 1969 Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. J. Mol. Biol. 40: 93–105.

Silva J. C., Loreto E. L. and Clark J. B. 2004 Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol. 6:57–71.

Smit A., Hubley R. and Green P. 1996. RepeatMasker Open-3.0. Available from: http://www.repeatmasker.org

Song S. U., Gerasimova T., Kurkulos M., Boeke J. D. and Corces V. C. 1994 An env-like protein encoded by a Drosophila retroelement: Evidence that Gypsy is an infectious retrovirus. Genes Dev.; 8:2046–2057.

Stoeckle B. C., Cowley D. E., Schaack Y., Macdonald III K. S., Sallenave R., et al. 2013 Microsatellites for North American species of *Triops* (Branchiopoda: Notostraca). J Crustacean Biol 33: 48–55.

Sun C., Shepard D. B., Chong R. A., Lopez A. J., Hall K., Castoe T. A., Feschotte C., Pollock D. D. and Mueller R. L. 2011. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol. Evol. 4, 168–183.

Temin H. M., 1980 Origin of retroviruses from cellular moveable genetic elements. Cell 21:599–600.

Thompson B. K. and Christensen S. M. 2011 Independently derived targeting of 28S rDNA by A- and D-clade R2 retrotransposons: Plasticity of integration mechanism. Mob Genet Elements. 1(1):29-37.

Tinti F. and Scali V., 1995. Hemiclonal, all-paternal and mosaic offspring of the hybridogenetics stick insect *Bacillus rossius-grandii grandii:* allozymic and cytological evidence. J. Exp. Zool., 273:149-159.

Valizadeh P. and Crease T.J. 2008. The association between breeding system and transposable element dynamics in *Daphnia pulex*. J Mol Evol 66: 643–654.

Vanschoenwinkel B., Pinceel T., Vanhove M. P. M., Denis C., Jocque M., Timms B. V., Brendonck L. 2012 Toward a Global Phylogeny of the “Living Fossil" Crustacean Order of the Notostraca. PLoS One. 2012; 7(4): e34998.

Van Valen L. 1973. A new evolutionary law. Evolutionary Theory 1: 1–30.

Vasil’eva L.A., Bubenshchikova E. V, and Ratner V. A. 1998. New evidence for the induction of mobile genetic element transpositions by severe heat shock. Genetika 34:1243–1250.

Velonà A., Luchetti A., Scanabissi F., Mantovani B. 2009 Genetic variability and reproductive modalities in European populations of *Triops cancriformis* (Crustacea, Branchiopoda, Notostraca). Ital J Zool 76: 366–375.

Wellauer P.K. and Dawid I.B. 1977. The structural organization of ribosomal DNA in Drosophila melanogaster. Cell 1977; 10:193-212

Wessler S.R. 1996. Turned on by stress. Plant retrotransposons. Curr. Biol. 6:959–961.

Wessler S.R., 2006. Transposable elements and the evolution of eukaryotic genome. PNAS 103:17600-1760.

Xiong Y. and T. H. Eickbush 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9:3353–3362.

Zaffagnini F., Trentini M. 1980. The distribution and reproduction of Triops cancriformis (Bosc) in Europe. Monitore Zoologico Italiano (N.S.), 14: 1-8.

Zhang X., Eickbush M. T. and Eickbush T. H. 2008. Role of recombination in the long-term retention of transposable elements in rRNA gene loci. Genetics 180: 1617–1626.

Zierold T., Montero-Pau J., Hänfling B., Gomez A. 2009 Sex ratio, reproductive mode and genetic diversity in *Triops cancriformis*. Freshw Biol 54: 1392–1405.

Zuccolo A., Sebastian A., Talag J., Yu Y., Kim H., Collura K., Kudrna D. and Winget R. A. 2007. Transposable element distribution, abundance and role in genome size variation in the genus Oryza. BMC Evol. Biol. 7, 152.