1. INTRODUCTION ... **3**
1.1 ABSTRACT ... 3
1.2 THE ORDER NOTOSTRACA (BRANCHIOPODA, CRUSTACEA, ARTHROPODA) ... 5
1.2.1 Morphology ... *6*
1.2.2 Ecology and distribution ... *7*
1.2.3 Reproductive biology ... *7*
1.2.4 Taxonomy and phylogeny ... *8*
1.2.5 Species of interest ... *9*
1.3 THE GENUS *BACILLUS* (INSECTA, PHASMATODEA, ARTHROPODA) ... 11
1.3.1 Morphology, ecology and distribution of the genus Bacillus ... *12*
1.3.2 Reproductive biology ... *13*
1.3.3 Taxonomy and phylogeny ... *14*
1.3.4 Species of interest ... *14*
1.4 REPETITIVE DNA ... 16
1.4.1 Ribosomal DNA: structure and evolution ... *16*
1.4.2 Transposable elements ... *18*
1.4.3 Transposable elements and reproductive strategies ... *20*
1.4.4 LINE transposable elements ... *21*
1.5. THE R ELEMENTS AND R2 ... 23
1.5.1 Structure ... *24*
1.5.2 Retrotranscription mechanism and activity studies ... *25*
1.5.3 Phylogeny ... *26*
1.6 MAIN AIMS ... 28
2. MATERIALS AND METHODS ... **29**
2.1 GENOMIC DNA ISOLATION ... 29
2.2 R2 ISOLATION, MOLECULAR CHARACTERIZATION AND PHYLOGENETIC ANALYSIS ... 29
2.2.1 Samples ... *29*
2.2.2 R2 detection/isolation ... *30*
2.2.3 R2 sequencing ... *31*
2.2.4 Phylogenetic analysis ... *34*
2.3 R2 ACTIVITY ... 34
2.3.1 Samples ... *34*
2.3.2 Truncation variant detection ... *39*
2.4 QUANTIFICATION OF RDNA UNITS AND R2 COPIES IN *T. CANCRIFORMIS* ... 40
2.5 28S RDNA SEQUENCE VARIATION ... 40
3. RESULTS ... **43**
3.1 THE R2 ELEMENT IN SPANISH *T. CANCRIFORMIS* ... 43
3.1.1 Structure ... *43*
3.1.2 Truncation analysis ... *43*
3.1.3 Sequence variability analysis of the 28S rDNA ... *46*
3.2 THE R2 ELEMENT IN ITALIAN AND AUSTRIAN *T. CANCRIFORMIS* ... 47
3.2.1 Structure ... *47*
3.3 THE R2 ELEMENTS IN *L. LUBBOCKI* ... 48
3.3.1 Structure ... *48*
3.3.2 Truncation analysis ... *49*
3.4 THE R2 ELEMENT IN *L. ARCTICUS* ... 50
3.4.1 Structure ... *50*
3.4.2 Truncation analysis ... *50*
3.5 THE R2 ELEMENTS IN *L. COUESII* ... 51
3.5.1 Structure ... *51*
3.6 THE R2 ELEMENT IN *B. ROSSIUS* ... 52
3.6.1 Structure ... *52*
3.6.2 Truncation analysis in field collected specimens ... *53*
3.6.3 Truncation analysis in laboratory crosses specimens ... *54*
3.7 THE R2 ELEMENT IN *B. GRANDII GRANDII* ... 55
3.7.1 Structure ... *55*
3.8 SUMMARY ... 56
3.9 PHYLOGENETIC ANALYSES ... 57

4. DISCUSSION **66**
4.1 R2 ELEMENTS STRUCTURE 66
4.2 28S VARIATION ANALYSIS IN *T. CANCRIFORMIS* 67
4.3 R2 PHYLOGENY 68
4.4 TRUNCATION ANALYSIS 70
5. CONCLUSIONS **74**
REFERENCES **75**
SUPPLEMENTARY MATERIALS **84**

1. Introduction

1.1 Abstract

Eukaryotic ribosomal DNA constitutes a multi gene family organized in a cluster called nucleolar organizer region (NOR); this region is composed usually by hundreds to thousands of tandemly repeated units. Ribosomal genes, being repeated sequences, evolve following the typical pattern of concerted evolution.

The autonomous retroelement R2 inserts in the ribosomal gene 28S, leading to defective 28S rDNA genes. R2 element, being a retrotransposon, performs its activity in the genome multiplying its copy number through a "copy and paste" mechanism called target primed reverse transcription. It consists in the retrotranscription of the element's mRNA into DNA, then the DNA is integrated in the target site. Since the retrotranscription can be interrupted, but the integration will be carried out anyway, truncated copies of the element will also be present in the genome. The study of these truncated variants is a tool to examine the activity of the element. R2 phylogeny appears, in general, not consistent with that of its hosts, except some cases (e.g. *Drosophila spp*. and *Reticulitermes spp.*); moreover R2 is absent in some species (*Fugu rubripes*, human, mouse, etc.), while other species have more R2 lineages in their genome (the turtle *Mauremys reevesii*, the Japanese beetle *Popilia japonica*, etc).

R2 elements here presented are isolated in 4 species of notostracan branchiopods and in two species of stick insects, whose reproductive strategies range from strict gonochorism to unisexuality. From sequencing data emerges that in *Triops cancriformis* (Spanish gonochoric population), in *Lepidurus arcticus* (two putatively unisexual populations from Iceland) and in *Bacillus rossius* (gonochoric population from Capalbio) the R2 elements are complete and encode functional proteins, reflecting the general features of this family of transposable elements. On the other hand, R2 from Italian and Austrian populations of *T. cancriformis* (respectively unisexual and hermaphroditic), *Lepidurus lubbocki* (two elements within the same Italian population, gonochoric but with unfunctional males) and *Bacillus grandii grandii* (gonochoric population from Ponte Manghisi) have sequences that encode incomplete or non-functional proteins in which it is possible to recognize only part of the characteristic domains. In *Lepidurus couesii* (Italian gonochoric populations) different elements were found as in *L. lubbocki*, and the sequencing is still in progress.

Two hypothesis are given to explain the inconsistency of R2/host phylogeny: vertical inheritance of the element followed by extinction/diversification or horizontal transmission. My data support previous study that state the vertical transmission as the most likely explanation; nevertheless horizontal transfer events can't be excluded.

I also studied the element's activity in Spanish populations of *T. cancriformis*, in *L. lubbocki*, in *L. arcticus* and in gonochoric and parthenogenetic populations of *B. rossius*. In gonochoric populations of *T. cancriformis* and *B. rossius* I found that each individual has its own private set of truncated variants. The situation is the opposite for the remaining hermaphroditic/parthenogenetic species and populations, all individuals sharing – in the so far analyzed samples - the majority of variants.

This situation is very interesting, because it isn't concordant with the Muller's ratchet theory that hypothesizes the parthenogenetic populations being either devoided of transposable elements or TEs overloaded. My data suggest a possible epigenetic mechanism that can block the retrotransposon activity, and in this way deleterious mutations don't accumulate.

1.2 The order Notostraca (Branchiopoda, Crustacea, Arthropoda)

Figure 1: Dorsal view of a tadpole shrimp.
Photo: Steve Jurvetson
Source: http://flickr.com/photos/jurvetson/33981006/

The order Notostraca comprises crustaceans characterized by a shield-shaped carapace covering the anterior region of the body; these animals, due to their carapace shape, resemble a tadpole: in fact they are also called tadpole shrimps (Figure 1).

The order consists of a single family with two genera: *Triops*, Schrank, 1803 and *Lepidurus*, Leach, 1816, that are immediately distinguishable for the presence in *Lepidurus* of a supra-anal plate, lacking in *Triops* individuals (Figure 2).

They are distributed worldwide, except Antarctica (Korn et al., 2006), and occur usually in natural temporary pools or rice fields. From eggs, a typical crustacean larva hatches: the nauplius. While larvae are planctonic, adults live digging in the mud with the head and swimming around. They are omnivorous, so they both filter the sediment and prey on various organisms.

These animals are well known examples of "living fossils" (King and Hanner 1998; Fisher, 1990; Longhurst, 1955), because fossil forms from the Triassic (200 Myr) are almost indistinguishable from the extant ones.

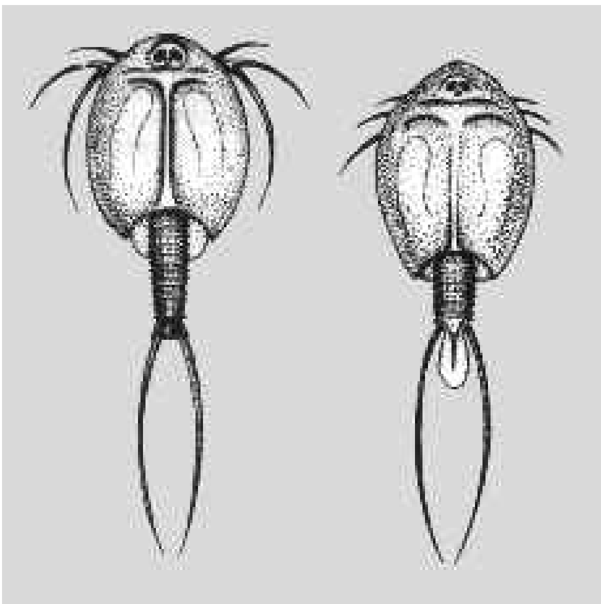

Figure 2: A *Triops* individual on the left and a *Lepidurus* individual on the right. Note the presence of the supra anal plate in *Lepidurus*.
http://www.sulinet.hu/eletestudomany/archiv/1999/9936/rakoka/rkoka.htm

Even if their morphology is very stable, individual characters, like the number of spines along the carina of the carapace or the number of abdominal segments, are usually very plastic. This fact, together with the lack of evident morphological discontinuities within the genera, makes the systematics, especially at the species level, very complicated, with species and subspecies often described on a single specimen taken as holotype.

1.2.1 Morphology

Notostracan body length ranges from 10 to 60 mm. The head (cephalon), the thorax (pereion) and part of the abdomen (pleon) are dorsally covered by a wide shield-shaped carapace with a middle carina, sometimes with spines, and a semicircular posterior emargination, the sulcus, in the hind edge of the shield, also with spines.

On the carapace, the odd naupliar eye and a pair of compound, sessile eyes followed by a dorsal organ, with sensorial function, can be found.

The cephalon appendages are the first and second pair of antennae, very reduced in size, a pair of mandibles and two pairs of maxillae.

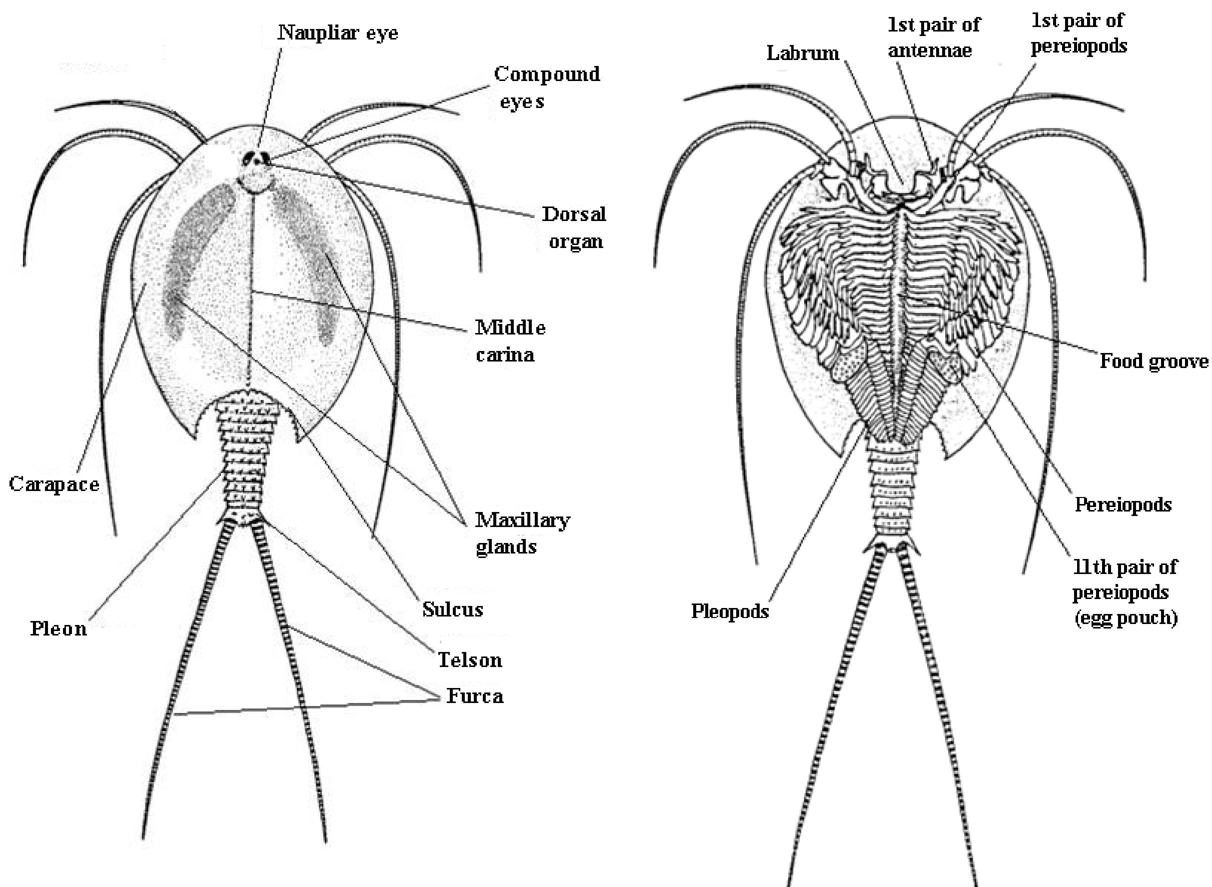

Figure 3: Dorsal and ventral schematic view of a *Triops cancriformis* individual. Modified from http://www.ub.edu/crba/practiques/zoologia/practica6/part1a.htm

The pereion comprises 11 thoracic segments each bringing a pair of appendages composed of lobes with diverse functions, in particular the endites of the first thoracic appendage are drawn out into filaments, of very variable length, with sensorial function. The pleon shows a variable number of segments, all bearing spines and a pair of appendages; the last abdominal segment (telson) bears, besides spines, a furca composed of two long and articulated uropods. In females, the eleventh pair

of limbs is different in structure from the others: the large outer lobe, which is used for swimming in other limbs, is modified into a round egg capsule, where the eggs are carried. In males, the eleventh pair of appendages does not differ from the others.

1.2.2 Ecology and distribution

Tadpole shrimps usually live near the ground of natural and artificial astatic pools, where they move with their ventral side down. However, the lack of oxigen can force them to swim upside-down with their gill-bearing legs close to the water surface. Notostraca are omnivorous; they dig around in the mud using the frontal part of their shield, searching not only for plankton but also for larger prey such as worms, chironomid larvae, small, dying or weak tadpoles, and even cannibalise freshly moulted, or smaller members of the same species. Notostraca survive in temporary pools all over the world, and are correspondingly short-lived. It seems that - in general - *Triops* species can survive only in waters which dry out regularly, while *Lepidurus* species show the tendency to live in pools that dry out less regularly, and *Lepidurus arcticus* is the only species that can live also in lakes; these ponds usually dry up during certain times of the year when there is no rainfall (Longhurst, 1955). Although adults die during these droughts, their resistant eggs can survive even for several years until the next rains fill up the pools again, allowing them to hatch. It is likely that resistant eggs are an adaptation to the ephemerality of the habitats and become efficient means of passive dispersal by the wind and migratory birds (Càceres and Soluk, 2002; Figuerola et al., 2005).

1.2.3 Reproductive biology

The morphological stasis of Notostraca contrasts with the extensive variation in reproductive biology: sexual reproduction ranges from gonochorism to hermaphroditism, from unisexuality to androdioecy, even in a single species as in the case of *T. cancriformis*. It's important to note that even when gonochoric reproduction takes place, sex-ratio may vary widely. For example in *T. cancriformis,* male frequency in different populations varies from 1% to 50% (Scanabissi et al., 2005). In *Triops granarius,* a gonochoric species, the sex ratios of its populations varies from 50: 50 to female bias or to male bias. The North American taxon *Triops newberryi* comprises bisexual populations with males at low frequencies and self-compatible hermaphrodites (Sassaman et al., 1997).

As far as species pertaining to the genus *Lepidurus* are concerned, they are generally indicated as gonochoric and/or hermaphrodites, but - as an example of the difficulties in determining reproductive strategies in these organisms - recent studies (Scanabissi and Mondini, 2002A; 2002B) carried out on an Italian population of *Lepidurus apus lubbocki* -the subspecies being indicated as

hermaphroditic by Longhurst (1955), then as gonochoric by Wingstrand (1978)- evidenced that males are not functional, so that the taxon appears to reproduce parthenogenetically.

For *L. arcticus* males are known, even if in very low proportion, only for Bear Island (Sømme, 1934); Longhurst hypothesized therefore a gonochoric and hermaphroditic reproduction, but unfortunately there aren't ultrastructural studies able to clarify this point.

Generally speaking, sex distribution is very difficult to settle without a hystological analysis, especially for the *Triops* genus; in fact, clear cut external sexual characters are lacking, so that only egg presence/absence can be observed, but this isn't resolving, given that hermaphrodites and females cannot be distinguished and it isn't applicable to juveniles.

1.2.4 Taxonomy and phylogeny

Notostraca exhibit a very high plasticity in external morphology, making species definition/recognition a difficult task on morphological grounds (Rogers, 2001). The absence of well-defined criteria allowed taxonomists to describe 'new species' based on morphological characters. Taxonomy and phylogeny of the group has been therefore the subject of wide discussion and frequent redescriptions (Linder, 1952; Longhurst, 1955; Lynch, 1966, 1970). More recently, the genetic and molecular studies of Sassaman et al. (1997), Suno-Uchi et al. (1997), King & Hanner (1998), Murugan et al. (2002), Mantovani et al. (2004, 2009) and Korn et al. (2006) put in light the existence of cryptic species.

The situation of the genus *Triops* seems to be in particular very complicated. Various Authors hypothesize the existence of cryptic or unrecognized species, because of their high level of genetic differentiation. For example, *T. granarius* is supposed to be composed at least by three cryptic species (Korn and Hundsdoerfer, 2006); the Authors, on the basis of 28S ribosomal gene and ribosomal mtDNA (12S and 16S genes) found that samples recognized as *T. granarius*, coming from different countries, show a genetic divergence of specific level. Murugan et al. (2002) studied *Triops* populations from Mexico using the mt12S gene, and found that the nominal (morphological) species *T. longicaudatus* is actually a mixture of several species.

Another example comes from the three classically recognized subspecies of the tadpole shrimp *T. cancriformis*. Korn et al. (2006) investigated their phylogenetic relationships using mitochondrial 16S and 12S rDNA sequences and they found out that the taxon, instead of being composed by three subspecies, can be divided into two distinct lineages. One lineage is constituted by *T. c. cancriformis* populations and samples from northern Spain that had been classified as *T. c. simplex* in the most recent literature. The second lineage comprises all populations of *T. c. mauritanicus* and northern African populations of *T. c. simplex* . Percentages of divergence are of specific level, as reported for other notostracan lineages and therefore Authors proposed to recognize them as two

species, *Triops cancriformis* and *Triops mauritanicus*. Finally, from molecular data it appears that the Old World species *T. cancriformis* may represent a lineage that is independent from other species in the genus *Triops* (Murugan et al., 2002; Mantovani et al., 2004).

The genus *Lepidurus* is classically known to occur in Europe with the species *L. apus*: the subspecies *L. apus lubbocki* is present in Italy and in all the Mediterranean countries, while in continental Europe the nominal subspecies occurs (Brtek and Thièry, 1995); through the analysis of mitochondrial markers (12S, 16S and COI) differences of specific rank between the two subspecies emerged (Mantovani et al., 2004; 2009). The Authors suggested to consider the two entities as distinct species, so that from now on I will use this distinction.

L. arcticus is another European species; the taxon shows a circumpolar distribution and it is capable of living also in permanent lakes. It's of very small dimensions with respect to other *Lepidurus* species, has a very small supra-anal plate and its eggs are purple instead of a bright red. In a recent study, Hessen et al. (2004), on the basis of molecular analysis performed on 12S rDNA from 48 circumpolar populations, recognized two main lineages of the species significantly differentiated also morphologically for the size of the supra-anal plate, but without a precise pattern of geographical distribution; nevertheless the possibility of intraspecific differentiation was suggested.

Lepidurus couesii is a North American and Asian species, its presence being reported also in Eastern Europe and Arabian peninsula (Brtek and Thièry, 1995). This species was completely unknown in the Italian fauna, but recently it has been recorded in southern Italy (Scanabissi et al., 2006). Longhurst, in his review (1955), considered *L. couesii* as synonym for *L. apus* and *L. packardi*; later Lynch (1972) considered the three taxa as different species. King and Hanner (1998) studied a fragment of the mt12S gene in four nominal species of *Lepidurus* from North America: *L. packardi*, *L. lemmoni*, *L. bilobatus* and *L. couesii*. The former were confirmed as valid species, while *L. couesii* appeared to comprise two genetically distinct lineages named *L. couesii*-1 (from California and Oregon) and *L. couesii*-2 (from Canada). Authors suggested that *L. couesii*-1 should retain the original name, because these samples seem to come from the same locality from which Linder (1952) examined his material. Finally, Rogers (2001) stated that *L. couesii*-1 is a different and new species: *Lepidurus cryptus*; therefore, in Rogers' view, *L. couesii*-2 corresponds to *L. couesii*.

1.2.5 Species of interest

During the PhD period I focused my work on four European Notostraca species: *T. cancriformis*, *L. lubbocki* (*sensu* Mantovani et al., 2009), *L. arcticus* and *L. couesii*. The four species are of particular interest because of their reproductive biology.

For *T. cancriformis* I have analyzed populations with different sex distribution: a gonochoric population sampled in two years (2004 and 2006) from Espolla (Spain), a hermaphroditic one from Marchegg (Austria) and a parthenogenetic population from Ferrara (Italy).
As far as *L. lubbocki* is concerned, I have analyzed the Italian population from Castel Porziano (Rome, Italy). This population was originally considered gonochoric because of the presence of males and females, but Scanabissi and Mondini (2002A and 2002B) demonstrated that males are sterile, so that the taxon should probably reproduce through parthenogenesis.
The populations of *L. arcticus* that I examined comes from two different localities in Iceland, and although they are composed only by egg-bearing individuals, it isn't clear if they are self-hermaphrodites or parthenogenetic; hystologic analyses that confirms their sexual status are in progress.
L. couesii population comes from Contrada Carracci (Lecce, Apulia) and it's gonochoric.
These species are of great interest because of their different reproductive modes, especially to investigate how they can influence the evolution and dynamics of repeated sequences and transposable elements.

1.3 The genus Bacillus *(Insecta, Phasmatodea, Arthropoda)*

Figura 4: A *Phobaeticus chani* individual. From http://el-blog.org/main-blog/17654-phobaeticus-chani-a-newly-discovered-stick-insect.html

The order Phasmatodea comprises phytophagous insects of middle-large size, or even huge like some tropical species (that can reach a length of 56.6 cm like the recently discovered *Phobaeticus chani* from Borneo, Fig. 4). These animals have exceptional mimetic capacities, that is their morphology and behaviour resemble the shape and the movements caused by the wind on branch and leaves: so they are also known as stick or leaf insects (Fig. 5), respectively.

The apterous forms are the rule; if winged, they aren't good flyer, preferring to walk with their long legs suited for arboreal-shrubby life.

Figure 6: *Bacillus rossius* eggs.

Figure 5: A leaf insect. Photo by Sandilya Theuerkauf, Wynaad

Their eggs (Fig. 6) are seed-shaped, and are laid singularly on the ground or attached in groups on various substrates like leafs, small breaks in trunks or in the ground; usually they have on the surface ornamental microstructures that are used as diagnostic characters for the systematics of the group.

Besides mimicry, these insects show other anti-predatory strategies like thanatosis, autotomy of limbs and emission of irritant substances.

The order Phasmatodea is present in all the equatorial, tropical and subtropical regions; the boundary of their distribution is placed around the 45° parallel of latitude N-S, even if some exceptions are given by England and South America.

In the Mediterranean area, 4 genera of stick insects are present: *Bacillus, Clonopsis, Leptynia and Ramulus*.

1.3.1 Morphology, ecology and distribution of the genus *Bacillus*

The genus *Bacillus* comprises holomediterranean, apterous species with nocturnal behaviour 6-10 cm long (Fig. 7), with a low mobility. The head is small, oval or oblong, with a pair of antennae more or less developed and a masticatory apparatus. Meso- and metathorax are well developed; the first abdominal segment is merged with metathorax forming the median segment.

Figure 5: A *Bacillus* stick insect on lentisk plant. http://fr.treknature.com/gallery/photo213080.htm

Their feeding plants are: blackberry bush (*Rubus spp*.), blackthorn (*Prunus spp*.), myrtle (*Myrtus communis*), dog rose (*Rosa canina*) and other Rosids like lentisk (*Pistacia lentiscus*); some species are monophagous, like *B. atticus* on lentisk.

Data let authors to consider that original feeding plants were lentisk and myrtle and afterwards some species adapted to blackberry, when it spread in semi-altered environments (Mantovani et al., 1991; Mantovani and Scali, 1993). Usually they live in natural/intact environments, but it is not so rare to find populations in semi-abandoned environments like blackberry bushes along railway roads or along motorways edges.

They are distributed mainly in the coastal regions and very rarely over 600 mt above sea level (Fig. 8).

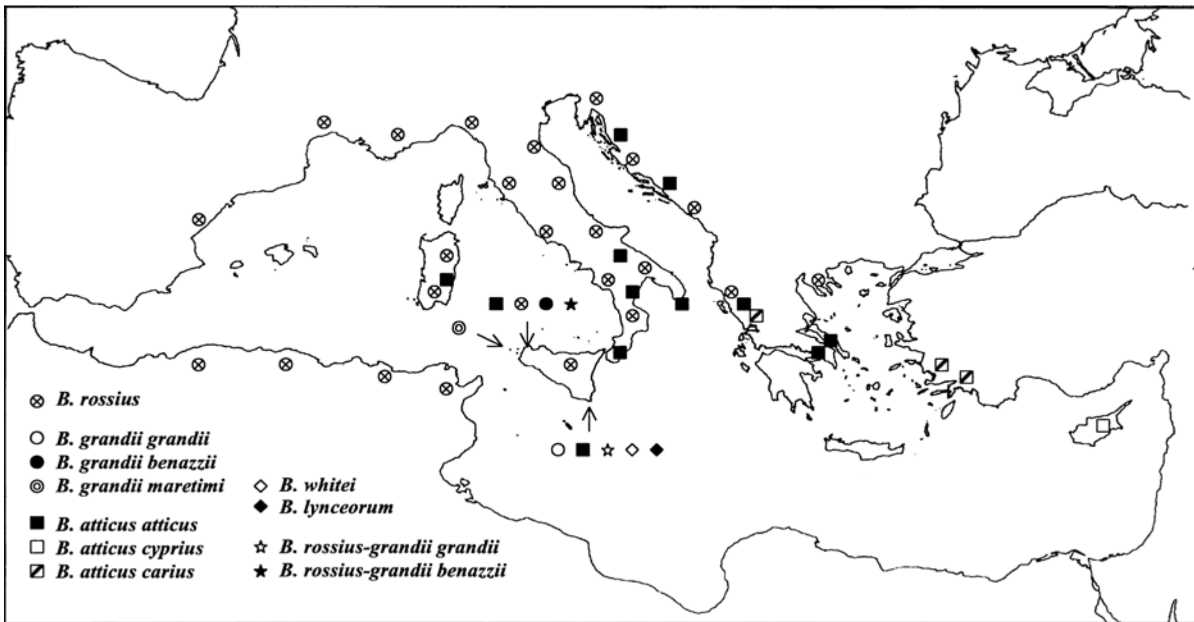

Figure 6: Map of the Mediterranean Basin showing the distribution of *Bacillus* parental (*B. rossius*, *B. grandii* and *B. atticus*) and hybrid taxa. From Mantovani *et al.*(1999).

1.3.2 Reproductive biology

Species pertaining to the genus *Bacillus* show a great variability and plasticity in reproductive strategies and represent one of the best known example of reticulate evolution (Fig 9). The genus comprises three parental species (*B. grandii* – gonochoric; *B. rossius* - facultative parthenogen; *B. atticus* - obligate parthenogen) that gave origin to a consistent group of hybrids: the diploid parthenogenon *B. whitei (B. rossius* X *B. grandii grandii*), the triploid parthenogenon *B. lynceorum (B. rossius* X *B. grandii grandii* X *B. atticus*) and the hybridogenetic *B. rossius-grandii*. Besides parthenogenesis and hybridogenesis, hybrids also show androgenetic reproduction, which produces even in nature individuals with a *B. grandii* nuclear genome and a mitochondrial *B. rossius* DNA (Mantovani and Scali, 1992; Mantovani et al., 1999; 2001).

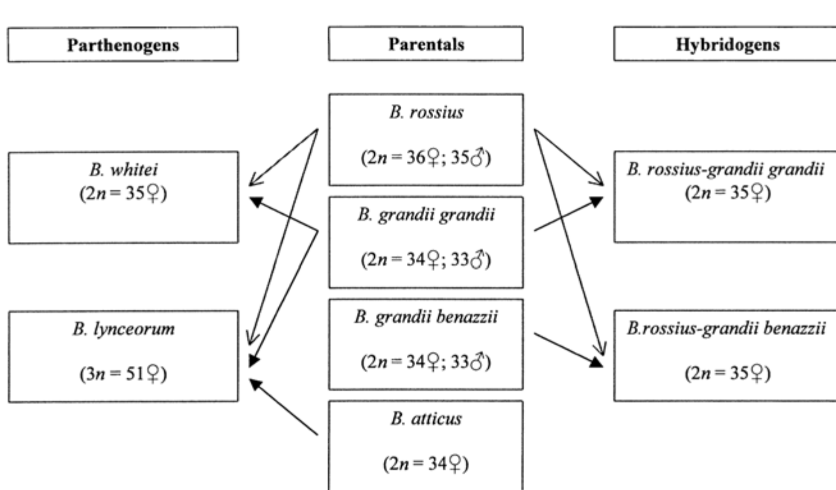

Figure 7: Schematic figure showing the origin of *Bacillus* hybrid taxa. From Mantovani *et al.* (1999).

1.3.3 Taxonomy and phylogeny

The three parental species are subdivided in subspecies on the basis of egg morphology (ootaxonomy), allozyme and karyological characterization (Scali and Marescalchi, 1987; Mantovani et al., 1999; Scali et al., 2003).

B. rossius is widespread in the western Mediterranean basin with eight subspecies; only two of them can be found in Italy: *B. rossius rossius*, on the middle-Northern Tyrrhenian coast and Sardinia (except Sarrabus), and *B. rossius redtenbacheri*, on middle-Southern Tyrrhenian coast, the whole Adriatic coast, Sicily and Sarrabus (Mantovani et al., 1991). The latter represents the maternal ancestor of all *Bacillus* hybrids (Mantovani et al., 1999).

B. grandii is endemic to the Sicilian region and it is differentiated in three subspecies: *B. grandii grandii* in the Iblean area with few relict populations; *B. grandii benazzii* in a very little zone in northwestern Sicily and on Levanzo Island and finally *B. grandii maretimi* that's endemic of the Marettimo Island (Mantovani et al., 1991).

B. atticus is present in Italy only with the nominal subspecies *B. atticus atticus*; *B. atticus carius* comprises the Greek and Turkish triploid demes and a diploid Turkish population, while *B. atticus cyprius* can be found on the island of Cyprus (Mantovani and Scali, 1993; Mantovani et al., 1999).

Phylogenetic relationships between taxa based on allozymes, satellite DNA and COII analyses are concordant providing evidences that *B. atticus* ssp. and *B. grandii* ssp. are more closely related to each other than with *B. rossius*, and surprisingly *B. grandii grandii* shows a greater affinity with *B. atticus* than with the other *B. grandii* subspecies. The analyses at the mitochondrial level demonstrated also that the maternal ancestor of all hybrids is *B. rossius*: in fact, in the phylogenetic tree based on COII gene, *B. whitei*, *B. lynceorum*, androgenetic *B. grandii grandii* and hybridogenetic *B. rossius-grandii benazzii* fall in the *B. rossius* clade (Mantovani et al., 2001; Scali et al., 2003).

1.3.4 Species of interest

I focused the attention on *Bacillus rossius*.

Five field collected samples were analyzed: two gonochoric populations from Anzio (Lazio) and Patti (Sicily), 3 sicilian parthenogens from Curcuraci, Massa San Nicola and Castanea delle Furie and one parthenogenetic population from Bertinoro (Fig. 10).

Figure 8: Map of sampling localities (indicated with red dots).

Then, in order to investigate the activity and transmission of retroelements, I analyzed also the offspring of two parthenogenetic females from Curcuraci and of two crosses between Curcuraci females and Anzio males.

The R2 element, sequenced in *B. rossius,* was also characterized in *B. grandii grandii* from Ponte Manghisi (Sicily) because I was interested in observing its behaviour in the element's phylogeny.

1.4 Repetitive DNA

It is well known that a substantial fraction of eukaryotic genomes is composed by repetitive DNA, namely homologous DNA sequences present in multiple copies. These sequences can be dispersed in the genome or tandemly arranged; interspersed repeats are those DNA elements that can be found in different sites in the genome (e.g. transposable elements -TE-), while tandem repeats are organized one following the other in a head-to-tail orientation (e.g. ribosomal DNA, satellite DNA, histone genes, etc.). In the next chapters, I will consider two types of repetitive sequences linked to thesis topic.

1.4.1 Ribosomal DNA: structure and evolution

The eukaryotic ribosomal DNA (rDNA) unit consists of three coding regions: 18S, 5.8S and 28S. Upstream the 18S, an external transcribed spacer (ETS) is present, while the 5.8S is separated from 18S and 28S by two internal transcribed spacers (ITS1 and ITS2). A non transcribed spacer (NTS), named also intergenic spacer (IGS), separates one unit from the other (Fig. 11).

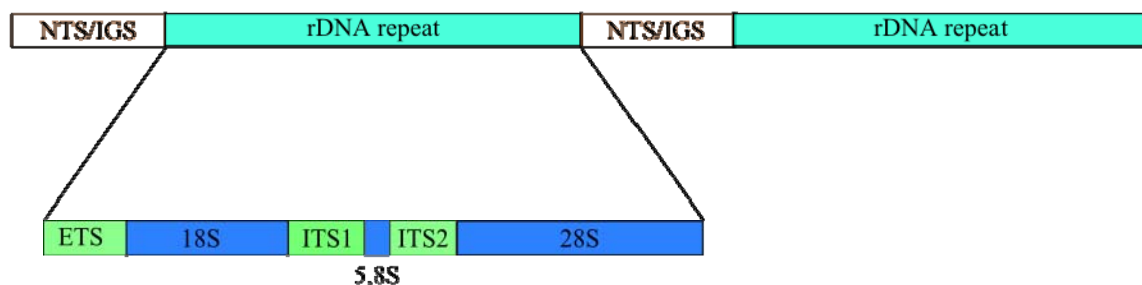

Figure 91: Schematic structure of ribosomal DNA units. Abbreviations are given in the text.

The number of rDNA units can vary from one (as in *Tetrahymena*) to hundreds of thousands per haploid genome (Weider et al., 2005). The number of units can vary even between individuals of the same species, as observed in frogs and fruit flies. Anyway, the total number of rDNA repeats seems to be in an excess in respect to what is needed for organism survival (Eickbush and Eickbush, 2007). Tandemly repeated rDNA units constitute the nucleolar organizer region (NOR), present on one (as in *D. melanogaster*) or more chromosomes (e.g. human, mouse, etc.). External and internal spacers and coding regions are transcribed as a single precursor that then will be spliced to obtain single rRNA subunits. The IGS contains the transcription and termination starting points. Beside this features, clusters of subrepeats are also present (Gorokhova et al., 2002; Luchetti et al., 2007); each subrepeat carries a gene promoter that act as transcription enhancer (Gorokhova et al., 2002; Hillis and Dixon, 1991). Exceptions to this trend are represented by the IGS of *Simulium sanctipauli* and *T. cancriformis*. In these organisms, the IGS subrepeats lack any

promoter sequence and it is, therefore, suggested that the presence itself of clusters of subrepeats act as enhancer for the rDNA transcription (Morales-Hojas et al., 2002; Luchetti et al., 2006).
Beside the ribosomal genes inside rDNA units, a 5S gene is present in eukaryotic genomes. The 5S genes can be interspersed in the genome (e.g. *Schizosaccaromyces pombe* and *Neurospora crassa,* Srivastava and Schlessinger, 1991) or organized in a cluster of tandemly repeated units; they are transcribed by an RNA polymerase III instead of a polymerase I.
The nucleotide sequence of the transcribed rDNA has different levels of conservation (Hillis and Dixon, 1991) among organisms: there are regions almost 100% conserved and regions poorly conserved. That's related to the secondary structure of the rRNA molecules: nucleotide sequences determining major features of secondary structure are usually highly conserved. Nevertheless, the secondary structure is maintained despite the perfect conservation of nucleotide sequence, because compensatory mutations occur (Hillis and Dixon, 1991).
Two main models are proposed to explain the evolution of multigene families: birth-and-death evolution and concerted evolution. The first model suggests that new variants arise by gene duplication, then some variants are maintained for a long time, while others are deleted or inactivated by deleterious mutations and overstay as pseudogenes (e.g. MHC in mammals) (Nei and Rooney, 2005). The model of concerted evolution was formulated by Dover (1982, 2002) after observing the pattern of "unexpected sequence homogeneity within and between individuals of a species" for a given gene family. It means that the repeated units don't evolve independently one from the other, but in a concerted manner; once a mutation arise in a unit, this new variant can spread to all the array (Fig.12).

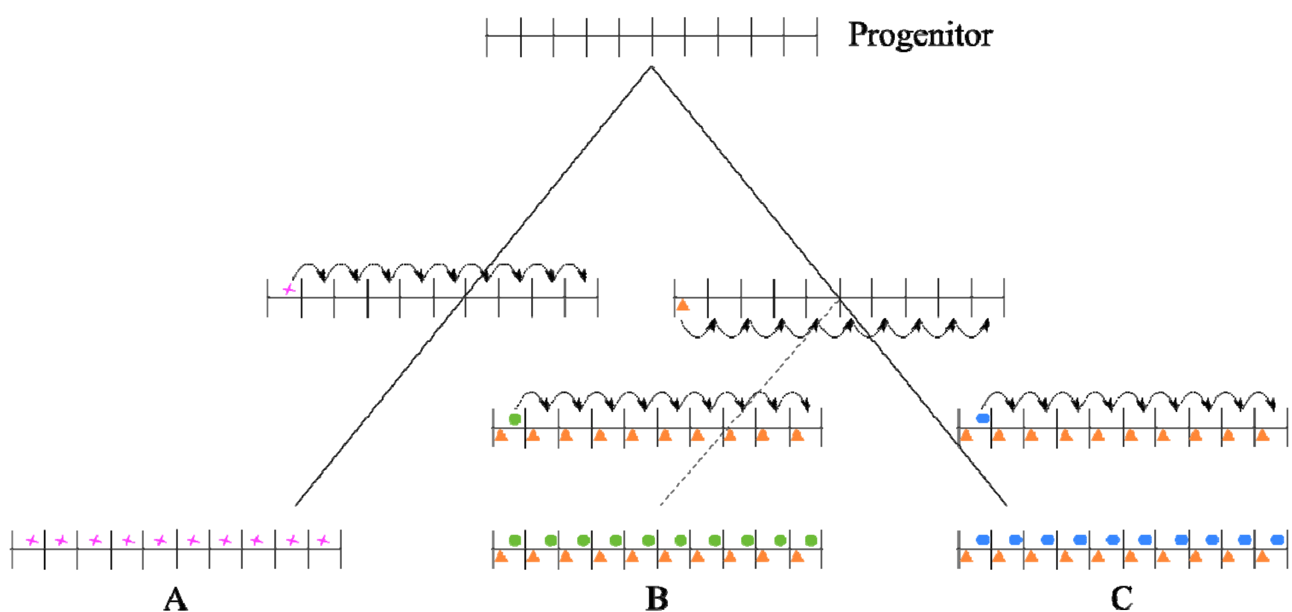

Figure 12: Graphic representation of the concerted evolution pattern of a multigene family from a common progenitor to three different derived taxa A, B, C. A mutation (represented by different geometrical symbols) arise in a unit and, after many generations, the new variant spread to all the array (modified from Dover, 1982).

The concept of sequence homogeneity can be extended to any evolutionary unit level, being this unit a chromosome, an individual, a population, a subspecies, a species and so on. On the contrary, if any member of a multigene family could be free to evolve independently, the sequence variability should be of the same magnitude among different evolutionary levels.

Sequence homogeneity is determined by the molecular drive, a process divided into two phases: gradual diffusion of a variant through the genome (homogenization) and through a population (fixation) (Dover, 1982; 2002). Homogenization is achieved through molecular turnover mechanisms such as unequal crossing over, gene conversion, slippage replication, eccDNA formation, amplification and reinsertion, transposition, retrotransposition. These mechanisms allow a mutation arisen in a unit (new variant) to be spread to all the units, while other mutations can be lost, leading to non-mendelian segregation patterns. The fixation of a variant through a population is achieved by the bisexual reproduction: in fact, thanks to the segregation of the homologous chromosomes and to panmixis, a variant can spread to the offspring and so through the entire population. Because the molecular drive process is stochastic, different variants can be fixed in different evolutionary units.

Ribosomal DNA shows a pattern of concerted evolution (Hillis and Dixon, 1991; Ganley and Kobayashi, 2007; Eickbush and Eickbush, 2007). The high sequence identity between units is achieved through unequal crossing over and gene conversion events (Stage and Eickbush, 2007); mutations occurring in the transcribed regions are subject to selection, due to rRNA secondary structure or transcription regulation, and are generally eliminated, while those arising in the non-trascribed regions are subject to a relaxed selective pressure and are more often fixed.

1.4.2 Transposable elements

Transposable elements (TE) constitute a large amount of the eukaryotic genomes, so that in some mammals and plants their percentage can reach 50% (e.g. human) or more (e.g. 85% in maize) (Kazazian, 2004; Wessler, 2006a, b). Thanks to whole genome sequencing projects, a lot of new types of TE are described since their discovery in maize by Barbara McClintock in 1940s (Wessler, 2006a). Because of their abundance and activity, TE continuously mould eukaryotic genomes and are considered one of the major forces underlying their evolution: in fact their presence at various sites in the genome can promote element-mediated rearrangements (e.g. recombination between TE

on non-homologous chromosomes) and alteration of genes expression (e.g. a gene interrupted by an element will not probably be transcribed).

Transposable elements are DNA fragments capable to move at various sites inside the genome (Kazazian, 2004). They are classically divided in two classes depending on whether the element has or has not an RNA intermediate: class 1 TE or retrotransposons -with RNA intermediates- and class 2 TE or DNA transposons (Fig. 13). Each class includes autonomous and non-autonomous elements, the latter deriving from the former. Autonomous elements encode proteins necessary for their activity, while the non-autonomous ones rely on the enzymatic machinery of the former for their activity because they retain the *cis* sequences necessary for the process. Class 1 retrotransposons are divided in two subclasses: LTR and non-LTR; LTR retrotransposons have long terminal repeats (LTRs) at both ends, while non-LTR retrotransposons lack LTRs and usually present a poly(A) tail at their 3' end (Wessler, 2006a).

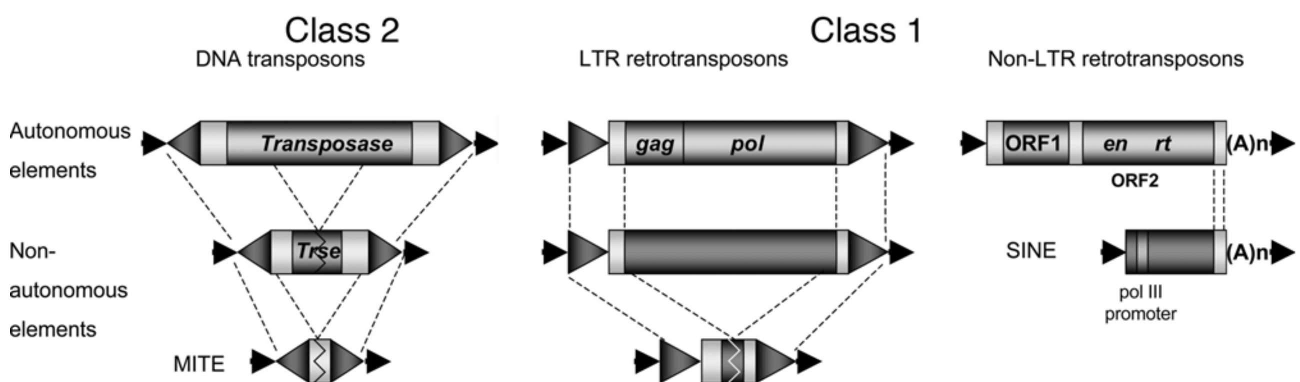

Figure 13: Classification of transposable elements as given by Wessler (2006b). Class 2 DNA transposons have a single open reading frame (ORF) coding for a transposase and terminal inverted repeats (large grey arrowheads); derived non-autonomous elements are called MITEs. Class 1 LTR retrotransposons have long terminal repeats in direct orientation (large grey arrowheads) and have two ORFs: *gag* -encodes for a capsid-like protein- and *pol* -encodes for reverse transcriptase-. Non-LTR retrotransposons have two ORFs: ORF 1 is a *gag*-like protein and ORF 2 encodes a protein with endonuclease and retrotranscriptase domains; derived non-autonomous elements are called SINEs and presents a polymerase III promoter; both have a poly(A) tail -(A)n-. Small black arrowheads indicate target site duplications.

Class 2 DNA transposons are excised from one site and reintegrated in another site in the genome through a *"cut and paste"* mechanism, while class 1 retrotransposons are transcribed into RNA,

then reverse transcribed and reintegrated into the genome through a *"copy and paste"* mechanism that enable the increase of their copy number. In both cases, after the transposition/retrotransposition process, short duplications of the target site can be produced.

1.4.3 Transposable elements and reproductive strategies

TEs are known to reduce the fitness of their host organism like most deleterious mutations, but, unlike other deleterious mutations, they are able of independent activity and therefore are first considered as genomic parasites (Dawkins, 1976; Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Kidwell and Lisch, 2001; Le Rouzic and Capy, 2005; Lee and Langley, 2010). This view of TE and their hosts fits with the Red Queen hypothesis: TEs are retained in the genomes, despite their deleterious effects, because they replicate faster than the host genome can stem them (Dolgin and Charlesworth, 2006) with deletion of copies, selection against over-inserted individuals and regulation of their activity (Hickey 1982; Charlesworth and Charlesworth 1983; Kaplan and Brookfield 1983; Le Rouzic and Capy, 2005). Coevolution between TEs and genomes is an arm's race in which sexual reproduction play an essential role. Outcrossing provides a means for TEs to spread virtually to all individuals in a population (Dolgin and Charlesworth, 2006).

This would not be true in unisexual populations. Under the assumption of Muller's ratchet theory, on the long term, asexual (unisexual) species/populations become extinct because of the over accumulation of deleterious mutations (Loewe and Lamatsch, 2008). TE insertions are considered as deleterious mutations, so it's expected that in unisexual organisms TE are absent. The only way by which a TE can spread in an asexual/unisexual population is by horizontal transmission. Examples of horizontal transmissions are well known (Sanchez-Gracia et al., 2005; Keeling and Palmer, 2008; Martirosyan et al., 2006; Kordiš and Gubenšek, 1999; Lampe et al., 2003), however it remains a rare event. Nevertheless, unisexual lineages arise from sexual progenitors whose genomes bear TEs and if element activity doesn't drive the lineage to extinction, there should be cellular mechanisms that limit them.

Sexual reproduction, in which two haploid genomes are combined, facilitates the spread of TE in a population, because it provides available niches in "virgin" genomes (Arkhipova, 2005). In fact, when an individual that has copies of a TE mates with an individual with no copies of that TE, the half-genome with no TE can represent an available niche for the element activity (Zhang and Eickbush, 2005). In this way a TE will rapidly become fixed in a population. But sexual reproduction made also possible to remove deleterious mutations in general, and TEs in particular, thanks to meiotic recombination and segregation (Arkhipova, 2005), because a TE present in a heterozygous state is transmitted to only one half of the offspring.

Actually, there are cases that contrast with the theory of Muller's ratchet. The best-documented exception to this rule is the Class Bdelloidea of the Phylum Rotifera, which has undergone successful radiative evolution in an ancient unisexual condition. These organisms have retroelements in a very low copy number (Gladyshev and Arkhipova, 2009) and generally they don't present TE insertions in the gene-rich regions, but in the telomeric/subtelomeric ones (Gladyshev and Arkhipova, 2010). It is also seen that they retain some DNA transposons, even if it isn't sure if they're active or not (Arkhipova and Meselson, 2000). Also the asexual protozoan *Giardia lamblia*, though lacking transposons and LTR retrotransposons, has two families of non-LTR retrotransposons that may be active (Arkhipova, 2005 and references therein). Schaack et al. (2010) studied DNA transposon dynamics in cyclical and obligate parthenogens of *Daphnia pulex* and found that they represents an exception to Muller's ratchet theory, because cyclical parthenogens had more transposons than the obligate ones, therefore corroborating data from Valizadeh and Crease (2008).

This fact confirms that sex facilitate the spread of TEs and suggests that presumably the host cells activate some type of suppression machinery/epigenetic regulation of the TE activity or even domestication (Dolgin and Charlesworth, 2006; Volff, 2006), in order to limit deleterious insertions. There are two main ways of TEs silencing: transcriptional gene silencing (TGS) and post transcriptional gene silencing (PTGS). The most studied TGS of TEs is the methylation of protein coding genes (e.g. in mammals), while PTGS mechanisms comprise RNA interference by dsRNA formation (e.g. in *Caenorhabditis elegans*) and post transcriptional RNA degradation (Okamoto and Hirochika, 2001; Kazazian, 2004 and references therein). TGS and PTGS aren't mutually exclusive and often work together to limit TEs activity.

1.4.4 LINE transposable elements

Long interspersed elements (LINEs) are autonomous non-LTR retrotransposons (see paragraph 4.2). Malik et al. (1999) identified 11 lineages of non-LTR TEs with different coding capacities; despite each lineage specific aspect, all non-LTR elements can be divided in two structural classes (Fig. 14).

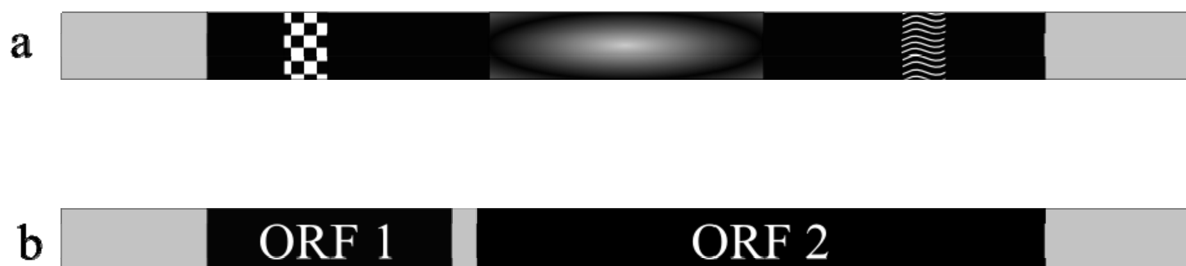

Figure 14: Comparison between the two structures of non-LTR elements. a) The first class has a single ORF (black box) encoding a DNA binding motif (checked box), a retrotranscriptase domain (faded box) and an endonuclease domain (wavy box). b) The second class has two distinct ORFs. ORF 1 encodes a *gag*-like protein; ORF 2 encodes for a protein with retrotranscriptase and endonuclease functions.

Both classes have untranslated regions (UTR) at their 5' and 3' termini, but while the first class has a single ORF, the second class has two ORFs. The ORF of the first class has a N-terminal domain that encodes DNA-binding motifs, a central retrotransposase domain (RT) and a C-terminal domain with DNA-binding and endonucleasic function (e.g. retrotransposon R2) (Eickbush, 2002; Christensen et al., 2006; Eickbush and Jamburuthugoda, 2008). These elements insert into highly conserved target sites, such as the ribosomal DNA genes of various animals (Eickbush, 2002; Eickbush and Eickbush, 2007). This feature seems to be an ancient characteristic of the group (Malik et al., 1999), therefore the first class TEs are proposed as the ancestors of all non-LTR retrotransposons (Malik et al., 1999; Burke et al., 2002).

The second class of non-LTR retrotransposons encodes two ORFs. The first ORF have functional similarity to the *gag* gene of retroviruses, and the protein has been shown to bind RNA (Martin and Bushman, 2001). The second ORF encodes a protein with endonuclease and retrotranscriptase domains; this endonuclease domain is called APE because of its similarity to apurinic–apyrimidinic endonucleases (APE) (Martin et al., 1995; Burke et al., 2002). The most extensively studied members of this group are the L1 elements of mammals and the R1 element of arthropods (Moran and Gilbert, 2002); while R1 inserts in a specific target site (28S ribosomal gene), L1 can be found in various sites in the genome (Kazazian, 2004 and references therein).

On the whole, the retrotranscriptase and the endonuclease domains seem to be the only well conserved features of the non-LTR elements (Malik et al., 1999; Burke et al., 2002; Eickbush and Malik, 2002).

Non-LTR retrotransposons are supposed to replicate through a mechanism called *target primed reverse transcription* or TPRT, in which the protein endonuclease domain nicks the target site and the element's transcript is used as a primer for the synthesis of the cDNA (Wessler, 2006a). The details of this mechanism depends on the single element's structural specific features and it was described in detail only for R2 (Christensen et al., 2006).

Phylogenetic analyses of a wide number of retrotranscriptase sequences highlighted that non-LTR TEs are monophyletic and are more closely related to group II mitochondrial introns and bacterial multicopy single-stranded DNA (msDNA) than to LTR elements (Xiong and Eickbush, 1990). The

most comprehensive phylogenetic analysis of non-LTR TEs, based on the RT domain, identifies 11 clades of elements identifiable through structural differences and dating back to the Cambrian period (Fig. 15) (Malik et al., 1999). The diversity of non-LTR retrotransposons and their wider distribution -in all eukaryotic organisms- suggest that they are the oldest group of retroelements (Xiong and Eickbush, 1988; Xiong and Eickbush, 1990; Eickbush, 1994).

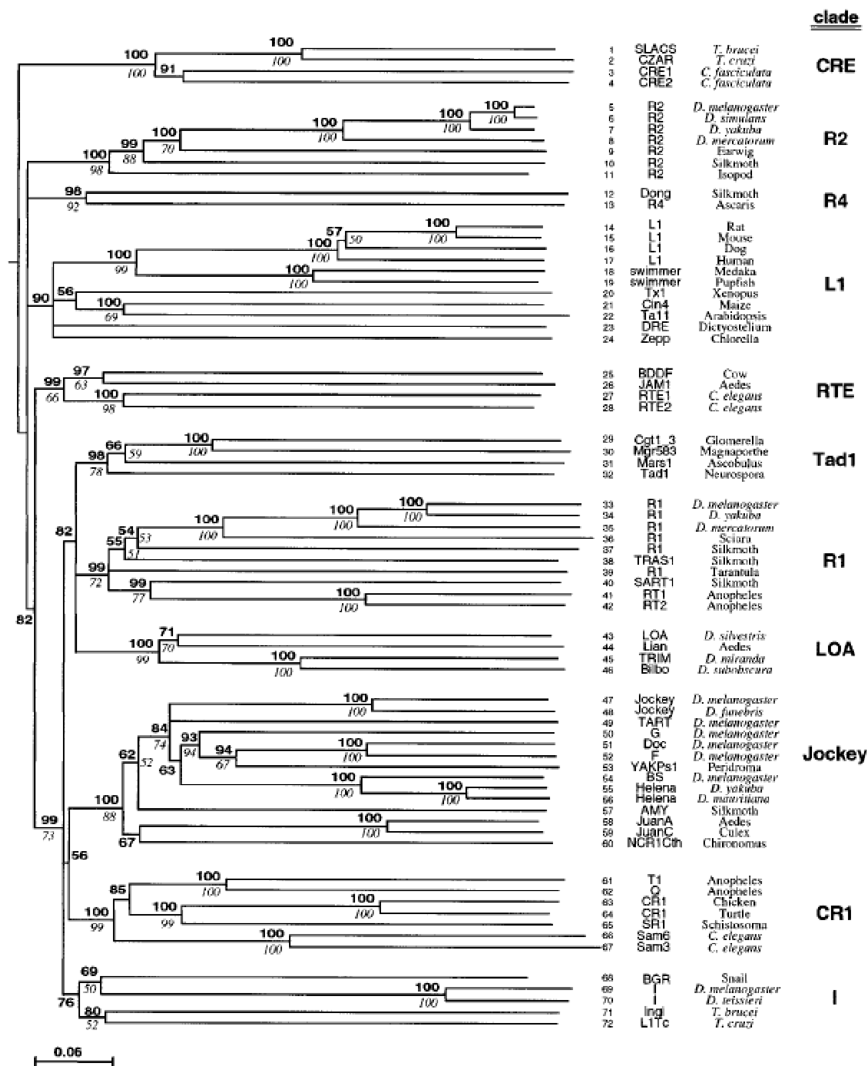

Figure 15: Phylogeny of non-LTR elements based on their entire RT domains. The phylogeny is a 50% consensus tree using the neighbor-joining method and rooted on the RT sequences of group II introns. Bold numbers next to each node indicate bootstrap values as percentages out of 1,000 replicates. Since Authors obtained a similar tree also by the maximum-parsimony method, bootstrap values are also given (numbers in italics below each branch point). From Malik et al., 1999.

1.5. The R elements and R2

The R elements are LINEs that inserts in the ribosomal genes. R4, R8, R5 and R9 are members of a common lineage found in many animal taxa and their structural organization is of the R2 type, the latter appearing the most diffused R element (Fig. 16; Kojima et al. 2006; Eickbush and Eickbush,

2007; Gladishev and Arkhipova, 2009). R6, R7, and RT represent a lineage related to R1; these elements can be found only in arthropods. All these elements are specific for ribosomal DNA genes (Fig. 16) and are rarely found outside the rDNA locus (Kojima and Fujiwara, 2003).

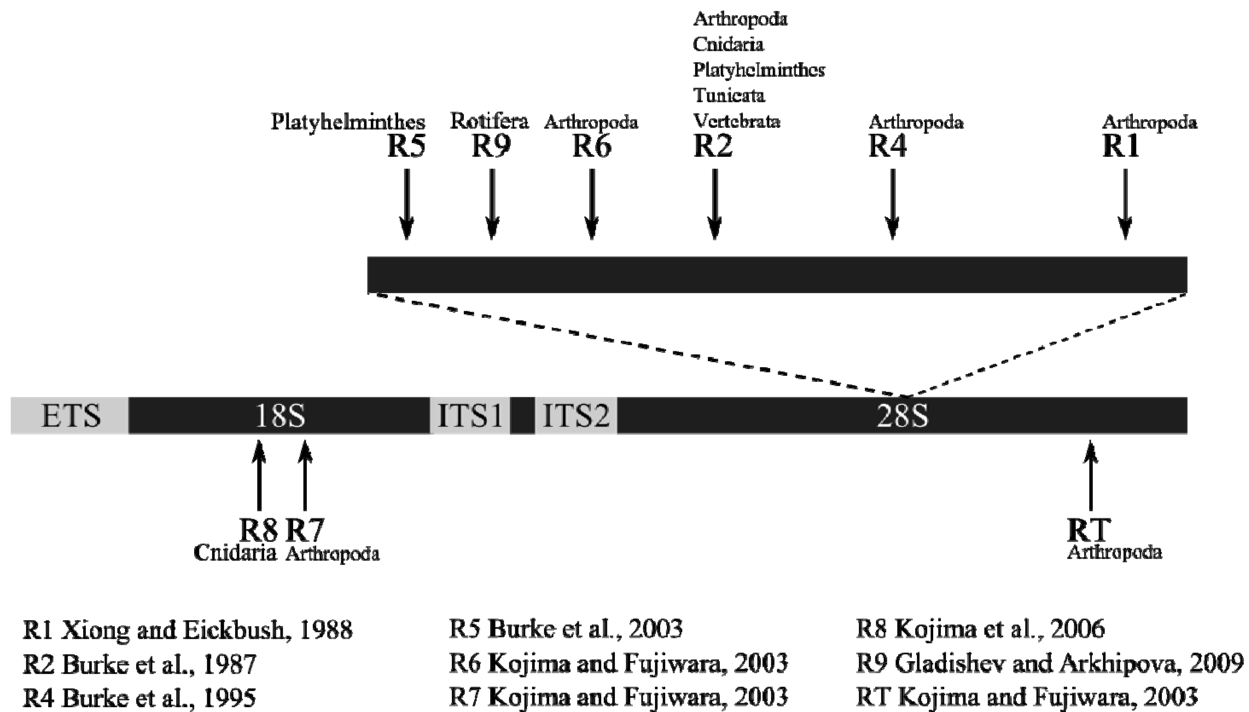

Figura 16: Scheme representing the R family of retrotransposons, their target genes and their hosts. At the bottom the references are given. Modified from Eickbush and Eickbush, 2007.

1.5.1 Structure

The nucleotidic sequence of the R2 element is not conserved among diverse organisms, but its structure comprises a 5'-end untranslated region (5'-UTR) followed by a single ORF and finally a 3'-end untranslated region (3'-UTR). The protein, as described above, is composed by an N-terminal DNA-binding domain, a central RT domain and a C-terminal DNA-binding and endonuclease domain (Christensen et al., 2006) (Fig. 17). These domains contain sets of amino acids that are usually well conserved: zinc-finger motifs, c-myb motif, CCHC motif and ENDO domain (Burke et al., 1999; Bunikis and Barbour, 2005).

R2 elements length ranges from 3.5 to 5 kb, with most of the length variation depending on the size of the 5'- and 3'- UTRs.

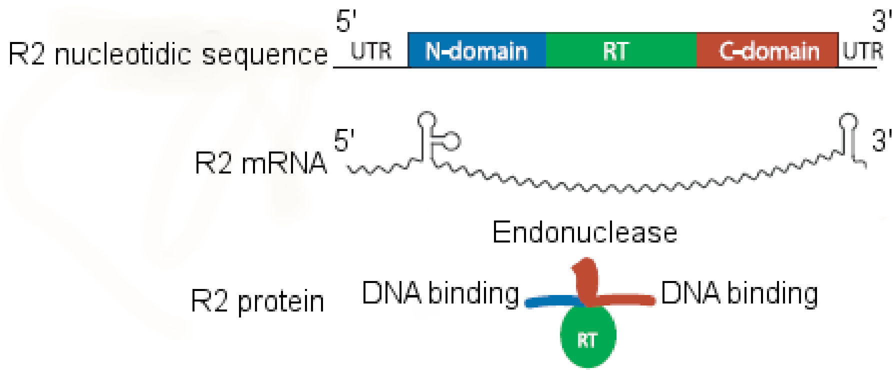

Figure 17: Model of the structure of the R2 element. Modified from Christensen et al., 2006.

1.5.2 Retrotranscription mechanism and activity studies

The mechanism by which R2 inserts in the host genome is called *target primed reverse transcription* or TPRT. First of all, as for all class 2 elements, the transcript moves to the cytoplasm where it is translated, and then both the mRNA and the protein return into the nucleus.

In the nucleus, two R2 proteins bind contemporaneously the 28S gene upstream and downstream the target site and the R2 transcript. Then the protein upstream cleaves the first target DNA strand – due to the endonuclease domain- release the 3' OH of the transcript and the RT domain, using the 3' OH released, begins the reverse transcription reaction. Once the 5' end of the transcript is released, the downstream subunit cleaves the second target DNA strand and begins the reverse transcription of the second R2 DNA strand (Fig. 18) (Christensen et al., 2006).

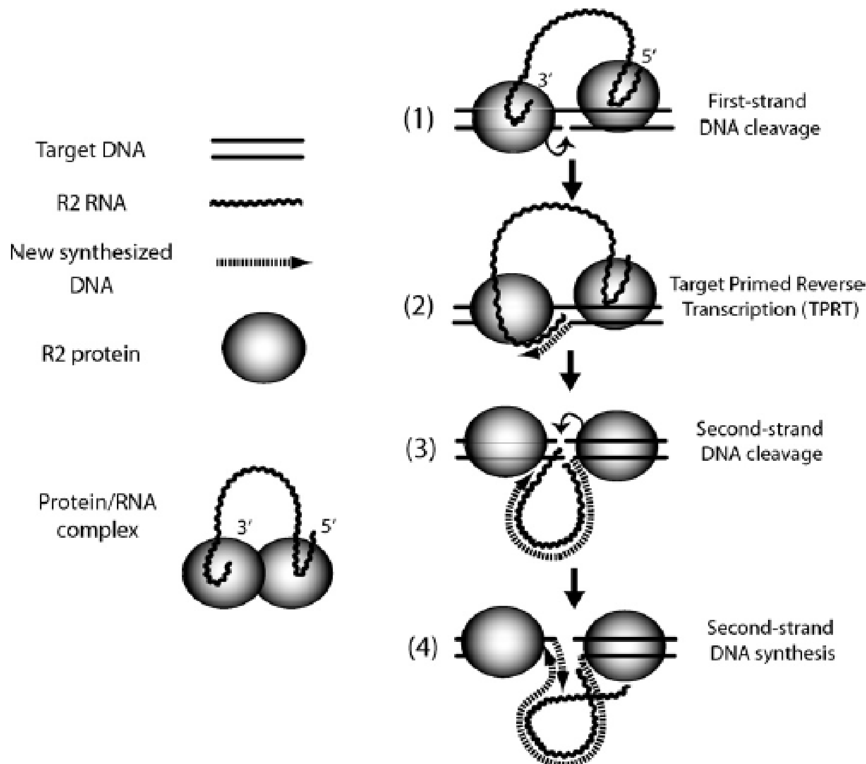

Figure 18: Model of R2 TPRT. From Eickbush and Jamburuthugoda, 2008.

The cleavage of the target site in two steps (first the top strand and then the bottom one) explains the presence of target site duplications/deletions of variable length.
Like other non-LTR retrotransposons, e.g. L1 in mammals, R2 presents a variable number of 5' truncated copies in the genome. That happens because the protein, during the retrotranscription, can dissociate from the mRNA before reaching the 5' end or, in alternative, cellular RNases can cleave the transcript. If one of these two things happens, a 5' truncated copy arises, because the incomplete element is integrated in the target gene anyway (Eickbush and Jamburuthugoda, 2008; Christensen et al., 2006).
The study of truncation variants is a tool to examine R2 activity: this aspect was deeply analyzed in laboratory stocks of *Drosophila spp*. R2 truncation analyses have been first conducted on *D. melanogaster* and *D. simulans*. In the former, variant distribution has been found, to some extent, well conserved, with ancestral-truncated variants being shared by individuals both within and between isofemale lines. However, some lines of *D. simulans* show decidedly higher R2 activity, producing less conserved truncation profiles (Pérez- Gonzalez and Eickbush, 2001; 2002; Pérez-Gonzalez et al., 2003; Zhang and Eickbush, 2005). A recent survey on natural populations of *D. simulans* showed a high turnover rate, each individual carrying a specific collection of R2 truncations (Zhou and Eickbush, 2009). The high incidence of R2 insertions in *D. simulans* is correlated with a high rate of variant elimination and a lower number of inserted 28S, explanable as due to its retrotransposition creating large deletions in adjacent rDNA units, thus eliminating a number of R2 variants (Zhang et al., 2008). R2 inserted 28S units aren't transcribed, so the rDNA unit is inactivated (Pérez- Gonzalez and Eickbush, 2001), although it's uncertain whether this inactivation is complete and whether it occurs in all tissues (Eickbush et al., 1997). Thereby, the selective pressure against non-functional rDNA units tends to eliminate the R2-inserted copies through the unequal DNA exchanges acting in the concerted evolution of the ribosomal locus (Eickbush, 2002; Eickbush and Eickbush, 2007). One of the effects of this turnover is the creation of new sites for the integration of new R2 copies. The percentage of inserted ribosomal genes within a genome can vary widely, but it's usually less than 50% (Eickbush et al., 1997; Bunikis and Barbour, 2005); that means that uniserted rDNA units sustain a sufficient level of rRNA synthesis, because deleterious fitness effects have only been noted in species with extremely high level of insertions (Eickbush et al., 1997).

1.5.3 Phylogeny

R2 elements' phylogeny represents an interesting matter for its long debated interpretation.

R2 was first identified in *D. melanogaster* (Burke et al., 1987) and then in other arthropod species (Burke et al., 1993; 1999) and it was considered to be vertically inherited since the raising of arthropods. Later, R2 was found in species pertaining to other animal phyla like Platyhelminthes, Echinodermata, Chordata and Cnidaria (Kojima et al., 2006; Kojima and Fujiwara, 2005) and the phylogeny of these elements was found to show several incongruences with that of their hosts. For example, R2 elements from *Drosophila spp.* are more closely related to those of earwig than they are to another dipteran (*Sciara coprophila*, fungus gnat); R2 from the crayfish *Procambarus clarkii* is closely related to that of the hagfish *Eptatretus burgeri* than to that of other crustaceans like the isopod *Porcellio scaber* and the tadpole shrimp *T. longicaudatus* (Kojima and Fujiwara, 2005). Moreover, some species, e.g. the jewel wasp *Nasonia vitripennis* and the Japanese beetle *P. japonica*, own different R2 elements that don't cluster together: in fact, their divergence ranges from 40 to 65% (Eickbush, 2002). On the other hand, R2 absence in taxa closely related to species harbouring it indicates that the extinction of this retrotransposon has occurred several times, at least during insect and vertebrate evolution (for example, in *Drosophila erecta*, *Drosophila orena*, *F. rubripes*, mouse, and human) (Eickbush et al., 1997; Kojima and Fujiwara, 2005). Two hypotheses have been put forward to explain this pattern: in the first, vertical inheritance of the element can be followed by lineage extinction/diversification in certain groups; the second hypothesis assumes the horizontal transfer of R2 between species. In a recent survey, the former has been shown as the most likely explanation (Kojima and Fujiwara, 2005), so the incongruence between host and R2 phylogeny can be explained, almost totally, by high rates of diversification of the element and not by horizontal transfer between species. This explanation is corroborated by the fact that the RT of R2 has a very low fidelity, estimated equal to that of HIV-1 virus (Eickbush and Jamburuthugoda, 2008). Obviously, the two hypotheses may not be mutually exclusive.

The complexity of terminal branch topology is counterbalanced by the essentiality of R2 deep nodes topology that follows a structural feature that is the number of zinc-finger motifs at the N-terminus (Kojima and Fujiwara, 2005). It would be interesting to investigate the factors affecting this feature and its evolutionary implications.

1.6 Main aims

My thesis work is based on the study of R2 dynamics in relation to reproductive strategies in non-model organisms. Notostracan crustaceans and stick insect species are of particular interest because of their reproductive biology, that ranges from gonochorism to parthenogenesis, even in populations of the same species. This pool of species and populations can constitute a model for the study of retrotransposons dynamics because it is possible to make comparisons among species/populations with canonical and non-canonical reproduction.

Studies published so far on R2 activity and evolutionary dynamics are conducted on the strict gonochoric model organisms *D. melanogaster* and *D. simulans* and nothing is known about its behaviour in low- or non-recombining genomes.

Beside the primary topic of R2 dynamic study, I was interested in observing two aspects related to the interaction element/host. The first was the effect of R2 insertion on the turnover of rDNA unit and how it can influence the concerted evolution. The second was the phylogenetic pattern of the element among strictly related species.

2. Materials and Methods

2.1 Genomic DNA isolation

Genomic DNA was isolated through one of the following methods: CTAB (Doyle and Doyle, 1987), phenol/chloroform protocol (Sambrook et al., 1989) or DNeasy Tissue Kit (Qiagen). These methods were applied depending on the quantity of tissue available and on the extraction yield: CTAB and DNeasy Tissue Kit are quick methods that enable to extract DNA from small tissue samples with a high yield in little time and are very useful with high numbers of samples. Phenol/Chloroform extraction permits to obtain very high quantity of DNA from both small and big tissue samples - even the entire animal - but in a longer time. CTAB and DNeasy Tissue Kit were used when the DNA was needed only for PCR amplification, while when the DNA was needed for other purposes, like digestions with restriction enzymes, the phenol/chloroform method was utilized (Tables 1, 5 and 6).

2.2 R2 isolation, molecular characterization and phylogenetic analysis

2.2.1 Samples

R2 has been isolated and characterized in four notostracans and two stick-insects species. Details on the samples are given in Table 1.

Table 1: Sampling years and localities, sample identification acronyms and DNA extraction method.

Species	Year	Site	Individual	Acronym	DNA
T. cancriformis	2004	Marchegg	Female 13	AU F13	CTAB
			Female 14	AU F14	
		Ferrara	Female 1	FRT F1	
			Female 3	FRT F3	
		Espolla	Male 1	04 M1	
			Female 22	04 F22	
L. lubbocki	2003	Castel Porziano	Male 1	T36 M1	Ph/Chlor
	2004		Female 1	T36 F1	
L. arcticus	2006	Holtavörðuheiði	Female 1	Hol 1	
		Arnavatnsheiði	Female 1	Arn 1	
L. couesii	2005	Contrada Carracci	Female 1	LcoCC ♀1	
			Male 1	LcoCC ♂1	
B. grandii grandii	2002	Ponte Manghisi	Male 52	Bgg PMA ♂52	
			Female 58	Bgg PMA ♀58	
B. rossius	2003	Capalbio	Male 1	Br Cap ♀1	
			Female 1	Br Cap ♂1	

2.2.2 R2 detection/isolation

One/two individuals for each species/site were utilized to check for the presence of the R2 element. Genomic DNA was PCR amplified with one of the forward degenerate primers described in Kojima and Fujiwara (2005), coupled with one of the 28S reverse primers (Table 2), located downstream of the element's insertion site. Reverse primers were designed with the online tool Primer3 (Rozen and Skaletsky, 2000), on the basis of tadpole shrimps and stick insects 28S gene sequences present in GenBank.

Table 2: Primers used to screen R2 presence in the analysed specimens.

Primer name	Sequence 5'>3'	Reference
R2IF1	AAGCARGGNGAYCCNCTNTC	Kojima and Fujiwara (2005)
R2IF2	GCYYTRGCGTTYGCNGAYGA	Kojima and Fujiwara (2005)
R2IIF1	GTNAARCARGGNGAYCCNCT	Kojima and Fujiwara (2005)
R2IIF2	CTNGCNTTYGCNGAYGAYYT	Kojima and Fujiwara (2005)
28S-R	TCCATTGCTGCGCGTCACTAATTAGATGAC	Designed for this work
28SB-R	CGTCTCCCACTTATGCTACACCTC	Mingazzini et al. (2010)
28SB-R_L	CGCCCTCCCACTTATGCTACACCTC	Designed for this work

Amplicons of the expected size (~2000 bp) were obtained in analyzed samples with different primers pairs as reported in Table 3 and they represent the 3' portion of the element.
PCR amplifications were performed in a 50 µl reaction mixture using the TaKaRa LA Taq with GC Buffer kit (TAKARA BIO Inc., Shiga, Japan), following the manufacturer's instructions. Thermal cycling was 94 °C for 5' , 94 °C for 30'', 48 °C for 30'', and 70 °C for 10' for 35 cycles; 15' at 72 °C as a final extension; TaKaRa kit was used for long amplifications, usually up to 2000 bp. Amplified PCR products were gel purified and cloned into a pGEM-T Easy vector (Promega, Madison, WY, USA). Sequencing was performed at Macrogen Inc. (Seoul, Korea). A couple of internal primers was designed on the basis of the sequences obtained; then the clones obtained previously were used as a template to make a nested PCR with the new primers. PCR amplifications were performed in a 50 µl reaction mixture using the Promega GoTaq Flexi kit, following the manufacturer instructions. Thermal cycling was 95 °C for 2' , 95 °C for 30'', 48 °C for 30'', and 72 °C for 1' 30'' for 35 cycles; 5' at 72 °C as a final extension; Promega kit was used for standard amplifications, usually less than 2000 bp.
Amplified PCR products were cleaned as described above and sequenced at Macrogen Inc.(Korea).

Table 3: Primer couples that gave an R2 amplification product for each species analyzed.

Species	Site	Primers pair
T. cancriformis	Marchegg	R2IF1 > 28SB-R
	Ferrara	R2IF1 > 28SB-R
	Espolla	R2IF1 > 28SB-R
L. lubbocki	Castel Porziano	R2IF2 > 28SB-R_L
L. arcticus	Holtavörðuheiði	R2IF2 > 28SB-R_L
	Arnavatnsheiði	R2IF2 > 28SB-R_L
L. couesii	Contrada Carracci	R2IF1 > 28SB-R_L
B. grandii grandii	Ponte Manghisi	R2IF1 > 28S-R
B. rossius	Capalbio	R2IF1 > 28SB-R

2.2.3 R2 sequencing

The complete sequence of R2 elements was obtained through the primer walking method: it consists in consecutive nested amplifications with internal primers since the sequence is completed as described above.

The starting amplicons were obtained amplifying genomic DNA of the specimens in Table 1 with the 28S-F2 forward primer coupled with an R2 species-specific reverse internal primer designed on the 3' portion of the sequence already completed. Only these amplicons were cloned as described above.

The primers were designed with the online tool Primer3 and are listed in Table 4.

PCR amplifications, amplified PCR products purification and sequencing were performed as previously described.

Sequences were edited and assembled using Chromas Pro version 1.42 (http://www.technelysium.com.au/ChromasPro.html) and MEGA4 (Tamura et al., 2007).

The open reading frame was found using the ORF Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf).

The protein conserved domains were found comparing the alignment of the newly characterized R2 amino-acid sequences to those found in Burke et al. (1999) and Bunikis and Barbour (2005).

Table 4: Primers designed for this study. Primers marked with an asterisk are published in Mingazzini et al. (2010). Forward primers are marked as F, while reverse primers are marked as R.

Species	Site	Primer name	Sequence 5'>3'	F/R
All species		28S-F2*	GTCAAAGTGAAGAAATTCAACGAAG	F
T. cancriformis	Espolla	DIN*	GGGTATTCAATTCTCGCATCTC	F
		DIN3*	AAGAGTCCTCAACAAAATTTTAAACCTACT	F
		DIN4*	TACAAAGAGCTCGTTAAAGATCAGC	F
		DIN5*	GGATAAGAGTAAGTGTTCTGTTTGTGG	F
		RIN*	GCAGGGAAAAAGAGGCATTAG	R
		RIN2*	GAACTCCAACTCTAAACAAGAGGTATCAG	R
		RIN3*	CTAGGTAGGAGTTAGTCAAGTCAAGCAG	R
		RIN4*	GATCTCTCAAATGAAGGAGTAGGTTTA	R
		RIN5*	CAATAATGTTGTCAAGTTTGTGTTCTA	R
		185'*	CCTGGTTGATCCTGCCAG	F
		18i*	TTTCTCAGGCTCCCTCTCCGGAATCGAACCCT	R
	Ferrara/Marchegg	R2ITDIN	ACGATGGTAAATCTCTTGCATACTC	F
		R2ITRIN	GAGATCAGGCCACGGTTATTTATC	R
		R2ITRIN2	CCAATAGATTTCTCTGCACTAAGTTTAAG	R
L. lubbocki	Castel Porziano	R2LDIN	GAGCCTAGAGACATGTATAAGTGGATT	F
		R2LDIN2	AATAAACCAGCTTGTTCCTAAATGTT	F
		R2LDIN3	GGAATTGTAGTGATGAAGTTCTGGA	F
		R2LDIN4	ATAAGCTAAAAGAGGCTAAGAAACCTG	F
		R2LDIN5	ACTCTAAGAAAAAGAGTAAATCAGTGG	F
		R2LRIN2	CAAAAGTCTCAGTCCAAAACTTCAT	R
		R2LRIN3	AATGAATTAGGTAGGTTCTCACAAAGA	R
		R2LRIN4	ATTATGGAAGAGTTAATGAAAGGAAGG	R
		R2LRIN5	AAGTAAGCTTATAACTTGCGCTTTGTA	R
		R2LRIN6	AACATAAGTTGTTTTCGCTTTACCA	R
		L2DIN	AGACAGGTGGGCTAGGGATT	F
		L2DIN2	GAGAAAGCCAATAATTCCTCTTCAT	F
		L2DIN3	GCTTTAAAATCAACTGTATGG	F
		L2RIN	GGAAACTAACGGGCATTCAC	R
		L2RIN2	CCGCTTCCGATAAAAGACAG	R

		L2RIN3	GAACCAGTACAGAGGAAATTGTGAT	R
		L2RIN4	AGTCGTCAGCAAACGCTAAAG	R
L. arcticus	Holtavörðuheiði/Arnavatnsheiði	LEPART-A-R2DIN	TGGGGATTAAAAACTTACGTACAAA	F
		LEPART-A-R2DIN2	ATCTATCCATTCTGAACTGCTCAAG	F
		LEPART-A-R2RIN	TGTCACTTAACAGATTTGCATTCAT	R
		LEPART-A-R2RIN2	GGTACACAGGTGAGTCTTCTTCAAT	R
		LEPART-A-R2RIN3	AACAACTATGGAGGAAATCGTGATG	R
		LEPART-A-R2RIN4	CCTGACTAGTGGGTTGAACTAGAAC	R
L. couesii	Contrada Carracci	LCO-C1_DIN	ATGGGTGAACTGAAAATAATGGATA	F
		LCO-C2_DIN	GGACTTGGGATTACGTGTTTTAGA	F
		LCO-D3_DIN	ATACCCCTTTGGGATTTTTCTACTC	F
		LCO-C1-RIN	AGGATATAGTTTTTGGAACGCATTA	R
		LCO-C2-RIN	AAGGACCTGGACAGAGGATTTAG	R
		LCO-D3-RIN	ATAAAGTTTTTGAAGCGCATTGAT	R
B. grandii grandii	Ponte Manghisi	BGG BDIN	AACGACTATCAGTTCGTTGAATAGGCT	F
		BGG BDIN2	AAACGGTAATCGGTCTAAACTCTTC	F
		BGG BDIN3	TATGATATTCAACATGGTTATTG	F
		BGG BRIN	ACATGATGTAATGTCTCAGTGCTGT	R
		BGG BRIN2	TATCCCACTCCATTCAATACAGTATCTC	R
		BGG BRIN3	CTCAAATCACAGCATATACATCAGG	R
		BGG BRIN4	ATGATAAGAGGGTATGGATAA	R
B. rossius	Capalbio	RR2DIN	AACGACTATCAGTTCGTTGAATAGG	F
		RR2DIN2	ACAGGACACTGGAGTCAATAAAGAG	F
		RR2DIN3	TGTGTGAAGTACTTACTAGATCG	F
		RR2RIN	CCATTCCATTCAATACAGTATCTCC	R
		RR2RIN2	AAGTCCTTTGGTGTAAAGGAATCTG	R
		RR2RIN4	GACCTGGTTCATCTTTTTAGCACT	R
		RR2RIN5	AAAGGACAGAGAGTGCCTTCAG	R
		RR2RIN6	ATGACAGAGTCAGGCTTCAGTAGAC	R

2.2.4 Phylogenetic analysis

The phylogenetic analyses were performed on amino acid sequences using the alignment of Mingazzini et al. (2010), to which the R2 sequences of termites published in Ghesini et al. (2011) were added: *Reticulitermes urbis* (R2Ru, Accession number: GU949554), *R. lucifugus* (R2Rl, GU949555), the two lineages of *Kalotermes flavicollis* (R2Kf1, GU949556; R2Kf2, GU949557), *R. grassei* (R2Rg, GU949558) and *R. balkanensis* (R2Rb, GU949559).

Amino-acid sequences were aligned with the MAFFT software online version (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/) using the E-INS-i algorithm (recommended for amino acid sequences with multiple conserved domains and long gaps) with BLOSUM62 scoring matrix for amino acids.

Neighbor Joining and Maximum Parsimony dendrograms were computed using PAUP* 4.0b10 (Swofford, 2003), with gaps treated as missing data; bootstrap values were obtained after 1000 replicates. The Bayesian phylogenetic tree was constructed using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). Monte Carlo Markov chains ran for 2 million generations, with trees sampled every 100 generations. In all analyses, the SLACS element (CAA34931) of *Trypanosoma brucei* (Aksoy et al., 1990) was used as outgroup.

2.3 R2 activity

R2 activity was studied through the analysis of 5' truncation patterns as described in Pérez-Gonzalez and Eickbush (2001). The 5' truncation variants were obtained by PCR amplification using a 28S gene primer upstream of the element insertion site and various R2 primers specific to locations within the element. Then amplification products were run on agarose gel and blotted on a nylon membrane and finally hybridized with specific probes. After the detection, positive bands were scored for each individual.

2.3.1 Samples

Samples used to screen R2 truncation patterns are listed in Table 5 and Table 6.

T. cancriformis were collected in Espolla (Spain); the same pond was sampled in 2004 and in 2006. Twenty individuals for each sampling were analyzed.

L. lubbocki samples come all from Castel Porziano (Rome, Italy) and are sampled in 3 years (2000, 2003 and 2004) from 4 ponds.

L. arcticus was sampled in 2006 in two different localities of Iceland: 11 egg bearing individuals from Holtavörðuheiði and 17 egg bearing individuals from Arnavatnsheiði.

As far as *B. rossius* is concerned, 68 individuals from 6 field- collected populations were analyzed: 19 from Curcuraci, 4 from Massa San Nicola, 4 from Castanea delle Furie and 2 from Bertinoro, 18 from Patti and 21 from Anzio. Furthermore, 40 laboratory-born specimens were analyzed: 10 females from both Female 2 and 3 from Curcuraci and 10 males from two crosses between Curcuraci females and Anzio males. Males were chosen from the crosses because they certainly represent hybrid descendants, while females can represent also parthenogenetic offspring. For all offspring 5 early-hatched individuals and 5 late-hatched individuals were chosen, to be sure that they represent distinct meiotic products.

Table 5: Sampling localities, years, samples identification acronyms and genomic DNA extraction method of field collected populations of the notostracan species examined.

Species	Year	Site	Individual	Acronym	DNA
T. cancriformis	2002	Marchegg	Female 13	AU F13	CTAB
			Female 14	AU F14	
	2003	Ferrara	Female 1	FRT F1	
			Female 3	FRT F3	
	2004	Espolla	Male 1	04 M1	
			Male 2	04 M2	
			Male 3	04 M3	Ph/Chlor
			Male 4	04 M4	
			Male 5	04 M5	
			Male 6	04 M6	
			Male 7	04 M7	
			Male 8	04 M8	
			Male 9	04 M9	
			Male 10	04 M10	
			Female 22	04 F22	CTAB
			Female 23	04 F23	
			Female 24	04 F24	
			Female 25	04 F25	
			Female 27	04 F27	
			Female 28	04 F28	
			Female 29	04 F29	
			Female 30	04 F30	
			Female 31	04 F31	
			Female 32	04 F32	
	2006		Male 1	06 M1	Ph/Chlor
			Male 2	06 M2	
			Male 3	06 M3	
			Male 4	06 M4	
			Male 5	06 M5	
			Male 6	06 M6	
			Male 7	06 M7	
			Male 8	06 M8	
			Male 9	06 M9	

			Male 10	06 M10	
			Female 11	06 F11	
			Female 12	06 F12	
			Female 13	06 F13	
			Female 14	06 F14	
			Female 15	06 F15	
			Female 16	06 F16	
			Female 17	06 F17	
			Female 18	06 F18	
			Female 19	06 F19	
			Female 20	06 F20	
L. lubbocki	2000	Castel Porziano	Male 2	RM M2	Ph/Chlor
	2003		Male 1	T36 M1	
			Female 3	R11 F3	
			Male 1	R11 M1	
			Female 4	PEP F4	
			Female 6	PEP F6	
			Male 13	PEP M13	
	2004		Female 1	T36 F1	
			Female 3	PEP F3	
			Male 1	PEP M1	
L. arcticus	2006	Holtavörðuheiði	Female 1	Hol 1	Ph/Chlor
			Female 10	Hol 10	CTAB
			Female 11	Hol 11	
			Female 12	Hol 12	
			Female 13	Hol 13	
			Female 14	Hol 14	
			Female 15	Hol 15	
			Female 16	Hol 16	
			Female 17	Hol 17	
			Female 18	Hol 18	
			Female 19	Hol 19	
		Arnavatnsheiði	Female 1	Arn 1	Ph/Chlor
			Female 2	Arn 2	CTAB
			Female 3	Arn 3	
			Female 4	Arn 4	
			Female 5	Arn 5	
			Female 6	Arn 6	
			Female 7	Arn 7	
			Female 8	Arn 8	
			Female 9	Arn 9	
			Female 10	Arn 10	
			Female 11	Arn 11	
			Female 12	Arn 12	
			Female 13	Arn 13	
			Female 14	Arn 14	
			Female 15	Arn 15	
			Female 16	Arn 16	
			Female 17	Arn 17	

Table 6: Sampling localities, years, samples identification acronyms and genomic DNA extraction method of *B. rossius* natural and laboratory born specimens examined.

Year	Site	Individual	Acronym	DNA
2009	Anzio	Female 1	An F1	Qiagen
		Female 2	An F2	
		Female 3	An F3	
		Female 4	An F4	
		Female 5	An F5	
		Female 18	An F18	
		Female 19	An F19	
		Female 20	An F20	
		Female 21	An F21	
		Female 22	An F22	
		Male 6	An M6	
		Male 7	An M7	
		Male 8	An M8	
		Male 9	An M9	
		Male 10	An M10	
		Male 11	An M11	
		Male 12	An M12	
		Male 13	An M13	
		Male 14	An M14	
		Male 15	An M15	
		Male 16	An M16	
	Patti	Female 1	Pa F1	
		Female 2	Pa F2	
		Female 3	Pa F3	
		Female 4	Pa F4	
		Female 12	Pa F12	
		Female 13	Pa F13	
		Female 14	Pa F14	
		Female 15	Pa F15	
		Female 16	Pa F16	
		Female 17	Pa F17	
		Male 5	Pa M5	
		Male 6	Pa M6	
		Male 7	Pa M7	
		Male 8	Pa M8	
		Male 9	Pa M9	
		Male 10	Pa M10	
		Male 11	Pa M11	
		Male 18	Pa M18	
	Curcuraci	Female 1	Cu F1	
		Female 2	Cu F2	
		Female 3	Cu F3	
		Female 4	Cu F4	
		Female 5	Cu F5	
		Female 6	Cu F6	
		Female 7	Cu F7	
		Female 8	Cu F8	

Year	Mother/Site	Offspring	Acronym	
		Female 9	Cu F9	
		Female 10	Cu F10	
		Female 11	Cu F11	
		Female 12	Cu F12	
		Female 13	Cu F13	
		Female 14	Cu F14	
		Female 15	Cu F15	
		Female 16	Cu F16	
		Female 17	Cu F17	
		Female 18	Cu F18	
		Female 19	Cu F19	
	Bertinoro	Female 1	Bert1	
		Female 2	Bert2	
	Castanea delle Furie	Female 1	CdF1	
		Female 2	CdF2	
		Female 3	CdF3	
		Female 4	CdF4	
	Massa San Nicola	Female 1	MSN1	
		Female 2	MSN2	
		Female 3	MSN3	
		Female 4	MSN4	
		Laboratory individuals		
Year	**Mother/Site**	**Offspring**	**Acronym**	
2010	♀2 Curcuraci	Female 1	BrCu♀2-F1	CTAB
		Female 5	BrCu♀2-F2	
		Female 6	BrCu♀2-F3	
		Female 7	BrCu♀2-F4	
		Female 10	BrCu♀2-F5	
		Female 164	BrCu♀2-F164	
		Female 169	BrCu♀2-F169	
		Female 170	BrCu♀2-F170	
		Female 172	BrCu♀2-F172	
		Female 173	BrCu♀2-F173	
	♀3 Curcuraci	Female 1	BrCu♀3-F1	
		Female 2	BrCu♀3-F2	
		Female 4	BrCu♀3-F4	
		Female 5	BrCu♀3-F5	
		Female 6	BrCu♀3-F6	
		Female 12*	BrCu♀3-F12*	
		Female 13*	BrCu♀3-F13*	
		Female 14*	BrCu♀3-F14*	
		Female 15*	BrCu♀3-F15*	
		Female 16*	BrCu♀3-F16*	
	Laboratory crosses			
	Parents/Site	**Offspring**		
	♀5 Curcuraci x ♂15 Anzio	Male 1	Cu ♀5 x An ♂15 M1	CTAB
		Male 4	Cu ♀5 x An ♂15 M4	
		Male 8	Cu ♀5 x An ♂15 M8	
		Male 9	Cu ♀5 x An ♂15 M9	
		Male 10	Cu ♀5 x An ♂15 M10	
		Male 133	Cu ♀5 x An ♂15 M133	

	Male 134	Cu ♀5 x An ♂15 M134
	Male 143	Cu ♀5 x An ♂15 M143
	Male 144	Cu ♀5 x An ♂15 M144
	Male 176	Cu ♀5 x An ♂15 M176
♀6 Curcuraci x ♂9 Anzio	Male 16	Cu ♀6 x An ♂9 M16
	Male 19	Cu ♀6 x An ♂9 M19
	Male 22	Cu ♀6 x An ♂9 M22
	Male 24	Cu ♀6 x An ♂9 M24
	Male 25	Cu ♀6 x An ♂9 M25
	Male 171	Cu ♀6 x An ♂9 M171
	Male 172	Cu ♀6 x An ♂9 M172
	Male 173	Cu ♀6 x An ♂9 M173
	Male 174	Cu ♀6 x An ♂9 M174
	Male 185	Cu ♀6 x An ♂9 M185

2.3.2 Truncation variant detection

Truncated element copies were obtained by PCR amplification (as described above) using the 28S-F2 primer, annealing 62 bp upstream of the element insertion site, coupled with various R2-specific reverse primers (Table 7) that anneal downstream of the insertion site.

For every primer pair, a specific probe was designed and used to hybridize the appropriate amplicon (Table7).

Table 7: Primers pairs and relative probes used for the truncation variant analysis.

Species	Primers pair	Length (bp)	Probe
T. cancriformis	28S-F2 > R2RIN2	3020	DIN > RIN2
	28S-F2 > R2RIN3	1899	DIN3 > RIN3
	28S-F2 > R2RIN4	1251	DIN4 > RIN4
	28S-F2 > R2RIN5	699	DIN5 > RIN5
L. lubbocki	28S-F2 > R2LRIN2	1899	LDIN4 > LRIN2
	28S-F2 > R2LRIN3	2126	LDIN > LRIN3
	28S-F2 > R2LRIN4	4186	LDIN3 > LRIN4
	28S-F2 > R2LRIN5	1472	LDIN2 > LRIN5
	28S-F2 > R2LRIN6	2864	LDIN5 > LRIN6
	28S-F2 > L2RIN	3324	L2DIN > L2RIN
	28S-F2 > L2RIN2	2006	L2DIN2 > L2RIN2
	28S-F2 > L2RIN3	1393	L2DIN2 > L2RIN3
	28S-F2 > L2RIN4	1880	L2DIN2 > L2RIN4
L. arcticus	28S-F2 > LEPART-A-R2RIN2	1995	28S-F2 > LEPART-A-R2RIN2
B. rossius	28S-F2 > R2R2RIN	1924	RR2DIN2 > RR2RIN
	28S-F2 > R2R2RIN2	1303	RR2DIN2 > RR2RIN2
	28S-F2 > R2R2RIN4	600	28S-F2 > R2R2RIN4
	28S-F2 > R2R2RIN5	2525	R2IF1 > RR2RIN5
	28S-F2 > R2R2RIN6	3033	RR2DIN > RR2RIN6

PCR products were separated on a 1,5% agarose gel in TAE 1X buffer; products longer than 2000 bp were run overnight, while products equal or shorter than 2000 bp were run for 6 hours.

DNA was blotted on a positively charged nylon membrane (GE Healthcare). Southern blots were performed washing the gel with Southern A solution two times for 20 minutes, followed by one wash with Southern B for 30 minutes (modified from Sambrook et al., 1989).
Membrane detection was performed using the DIG High Prime DNA Labeling and Detection Starter Kit I and the CDP-Star (Roche Diagnostics GmbH, Mannheim, Germany) for *T. cancriformis* analysis, while for all the other species the AlkPhos Direct Labelling Reagents with CDP-Star detection (GE Healthcare) was used.
The images were analyzed using Total Lab100 software (Nonlinear Dynamics, Ltd., Newcastle on Tyne, UK); through software evaluations, bands belonging to different individuals were considered the same truncation variant for differences up to ±10 bp.
For R2 of *T. cancriformis* from Marchegg and Ferrara the amplicon 28SF2>R2ITRIN was run and blotted (as described above) and hybridized with the probe R2ITDIN>R2ITRIN.
Statistic analyses on the presence/absence matrix of truncated variants were performed with the free softwares Famd (Schlüter and Harris, 2006) and Past (Hammer et al., 2001).

2.4 Quantification of rDNA units and R2 copies in T. cancriformis

Quantification of rDNA units and R2 copies within the *T. cancriformis* genome was performed through dot blot analysis. Genomic DNA was spotted onto positively charged nylon membranes in a series of dilutions (2000-15.6 ng); probe lanes had dilutions ranging from 5 to 0.04 ng for the 18S probe, and from 0.1 to 0.00078 ng for the R2 probe. To quantify the percentage of rDNA units, the blotted membrane was hybridized with a 400 bp long 18S probe obtained using primers 18-5’>18i. To score the percentage of R2-inserted units, the filter was hybridized with a 1309 bp probe specific for R2 (primer pair DIN>RIN). Hybridizations were performed under highly stringent conditions, with the final wash at 65 °C in 0.1X SSC, 0.1% SDS. Probe labelling and blot detection were performed using the DIG High Prime DNA Labelling and Detection Starter Kit I and the CDP-Star. Images were analyzed using ImageJ (Rasband, 1997–2007).

2.5 28S rDNA sequence variation

The nucleotide variability of the 28S genes harbouring or lacking R2 was analyzed through the amplification, cloning, and sequencing (as above described) of two regions extending from the R2 insertion site to 738 bp upstream and 810 bp downstream (Figure 19).

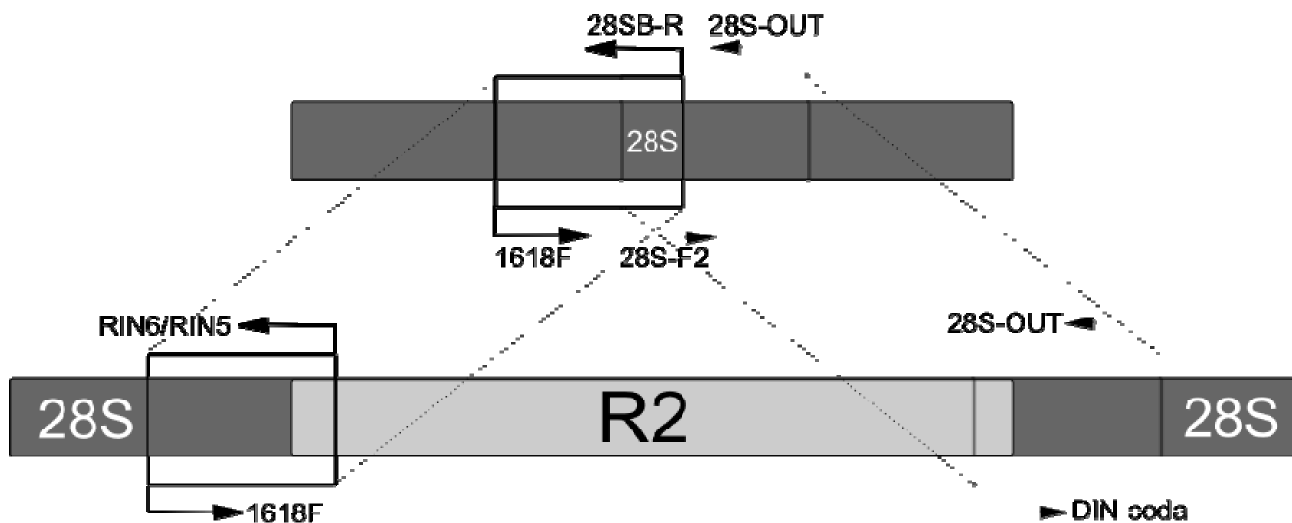

Figure 19: Scheme of the 28S region analyzed with primers name (Mingazzini et al., 2010).

Table 8: Primers used to amplify the 28S regions of interest. F stands for forward and R stands for reverse. RIN5 was used when RIN6, that is more internal, failed to amplify the target region. These primers are published in Mingazzini et al. (2010).

Primer name	Sequence 5'>3'	F/R
28S-F2	GTCAAAGTGAAGAAATTCAACGAAG	F
1618F	GAAAGGGAATCCGGTTCCCATTCC	F
DIN-coda	AAGTGGGAAGTGTTTTCAATGTACT	F
RIN5	CAATAATGTTGTCAAGTTTGTGTTCTA	R
RIN6	CACTGATAAATCAGCTTACCCAGTCT	R
28S-OUT	TTCAGGTATAATCAGACGGACGTAG	R
28SB-R	CGTCTCCCACTTATGCTACACCTC	R

Table 9: Primers pairs corresponding to each region amplified.

	Upstream region	Downstream region
W/O R2	1618F>28SBR	28S-F2>28S-OUT
W R2	1618F>RIN5/RIN6	R2DINcoda>28SOUT

The analysis was performed on eight individuals of the 2004 sample (M2, M3, M4, M6, F24, F25, F27, and F32); three of these specimens lack the R2 full-length element (M4, M6, F32).

Proportions of nucleotide differences (calculated as mean p-distances within each individual, p-D) and gene diversities (H) were calculated for both 28S regions; each value was taken as data point for further elaborations. Two-tailed Student t-tests, with equal variance, were computed to assess the significance of differences among the scored variability values assuming that clones harbouring

R2 are more variable than those lacking the insertion. A further comparison was performed between clones belonging to the individuals with the complete element (M2, M3, F24, F25, and F27) and individuals without the complete element (M4, M6, and F32), both for rDNA units harbouring R2 and lacking R2. Finally, a test for selection was performed on both single and pooled datasets using the Tajima's D parameter.

3. Results

3.1 The R2 element in Spanish T. cancriformis

3.1.1 Structure

To construct the complete sequence, six clones containing the whole insertion site at the 5' terminus (5'-TTAAGG↓TAGC-3'; Burke et al., 1999; Kojima and Fujiwara, 2005) were first considered; 5 clones showed deletions of the 28S gene ranging from 4 to 10 bp. The sequence of the full-length R2 element (R2Tc Spain) is a consensus of 5 sequenced fragments obtained by primer walking. It is 3583 bp long (GenBank A.N.: EU854578) and exhibits an A+T content equal to 53%. The R2 5' end showed a poly-T run of five nucleotides in all analyzed clones while at the 3' end a poly-A tail occurs, this being another common feature of R2 mobile elements (Burke et al., 1999; Kojima and Fujiwara, 2005). The R2 sequence contains a 3093 bp long ORF, located between nucleotide 177 and nucleotide 3272, coding for 1031 amino acids. The ORF is characterized by a reverse transcriptase domain and an endonuclease domain. Moreover, it exhibits a single 5'zinc-finger motif (Figure 20).

Figure 20: Schematic representation of *T. cancriformis* R2 element.

In *T. cancriformis*, R2 occurs at low copy number: only 0.54–5.3% of rDNA units (that constitute the 0.1% of the genome) have R2 insertions.

3.1.2 Truncation analysis

The results of Southern blots for R2 truncation patterns are summarized in Table 10 and Figure 21 (both from Mingazzini et al., 2010). In the 2004 sample, a total of 318 truncations were detected, ranging from 4 to 28 per individual. In the 2006 sample, 239 variants were scored, ranging from 2 to 20 per individual. Generally speaking, a wide range of truncation profiles has been scored, as each specimen shows its own set of truncations. No ancestral variants have been detected: indeed, there is not any truncation variant present in all individuals; the most widespread one is found in 10 and 7 individuals of 2004 and 2006 collections, respectively. Moreover, six individuals, from both 2004 and 2006 samples (Table 10; Figure 21), presented a set of R2 truncations (ranging from 2 to 20), but did not have the complete element.

Table 10: List of *T. cancriformis* samples used in the analyses and summary of truncation variants. From Mingazzini et al., 2010.

Year	*Sex*	*Alias*	*Complete R2*	*Truncations number*	*Total*	*Mean*
2004	Male 1	04 M1	+	28		
2004	Male 2	04 M2	+	20		
2004	Male 3	04 M3	+	13		
2004	Male 4	04 M4	−	11		
2004	Male 5	04 M5	+	17		
2004	Male 6	04 M6	−	6		
2004	Male 7	04 M7	+	13		
2004	Male 8	04 M8	+	16		
2004	Male 9	04 M9	+	5		
2004	Male 10	04 M10	+	4		
2004	Female 22	04 F22	+	22		
2004	Female 23	04 F23	+	17		
2004	Female 24	04 F24	+	16		
2004	Female 25	04 F25	+	19		
2004	Female 27	04 F27	+	25		
2004	Female 28	04 F28	+	17		
2004	Female 29	04 F29	+	17		
2004	Female 30	04 F30	+	18		
2004	Female 31	04 F31	+	20		
2004	Female 32	04 F32	−	14	318	15.9
2006	Male 1	06 M1	+	11		
2006	Male 2	06 M2	+	3		
2006	Male 3	06 M3	+	16		
2006	Male 4	06 M4	+	12		
2006	Male 5	06 M5	+	8		
2006	Male 6	06 M6	+	18		
2006	Male 7	06 M7	+	18		
2006	Male 8	06 M8	+	10		
2006	Male 9	06 M9	−	20		
2006	Male 10	06 M10	−	20		
2006	Female 11	06 F11	+	6		
2006	Female 12	06 F12	+	10		
2006	Female 13	06 F13	+	17		
2006	Female 14	06 F14	+	6		
2006	Female 15	06 F15	+	16		
2006	Female 16	06 F16	+	9		
2006	Female 17	06 F17	−	2		
2006	Female 18	06 F18	+	20		
2006	Female 19	06 F19	+	10		
2006	Female 20	06 F20	+	7	239	11.95
					557	13.93

+/− indicates the presence/absence of the complete element.

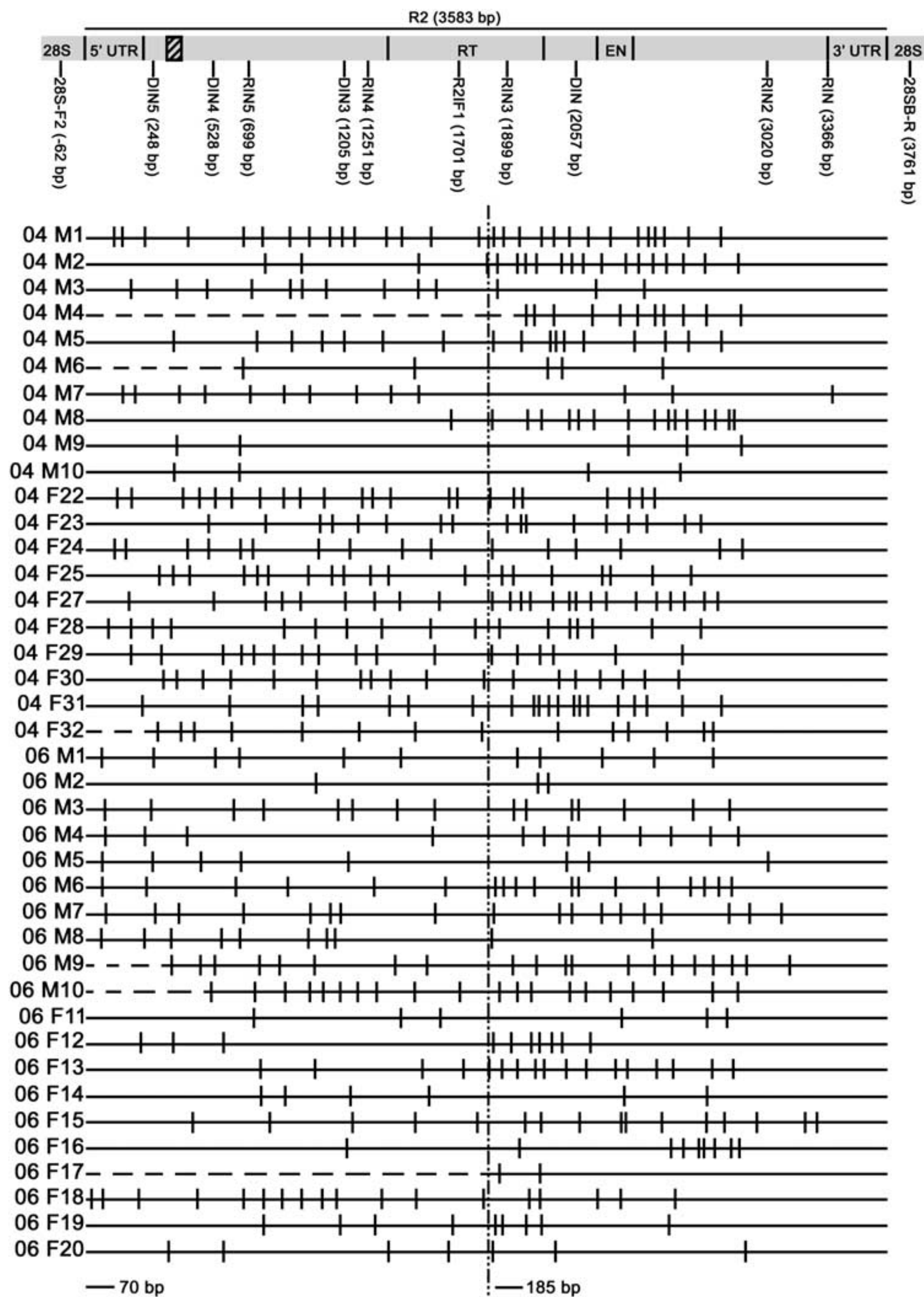

Figure 21: Graphic representation of R2 truncation variants distribution (shown by solid vertical lines) in the 40 *T. cancriformis* individuals. Dashed horizontal lines represent sequences missing in all elements from that individual; the dotted vertical line represents a change in the scale of the x axis. A diagram of the R2 element and the primer names/positions with respect to the insertion site are shown at the top. Pattern box indicates the zinc-finger motif; RT, retrotranscriptase domain; EN, endonuclease domain. From Mingazzini et al., 2010.

3.1.3 Sequence variability analysis of the 28S rDNA

The nucleotide variability of the 28S gene was studied using two sub-samples: the first one comprised individuals harbouring both complete and truncated R2 copies (individuals M2, M3, F24, F25, and F27), whereas the second sub-sample comprised individuals only with truncation variants, but lacking the complete element (individuals M4, M6, and F32; Table 11 from Mingazzini et al., 2010).

For each specimen, sequencing was performed for both R2-inserted and R2-uninserted 28S, upstream and downstream of the insertion site (Figure 19). From 6 to10 sequences per individual were obtained for 28S rDNAs with or without R2 insertions, for both upstream and downstream regions.

A total of 147 28S sequences 738 bp long were obtained for the upstream region: 76 carrying the element (hence R2+) and 71 without it (hence R2-). In the R2+ dataset, 110 polymorphic sites were found resulting in 56 alleles; the mean sequence diversity within individual varies widely, from 0.0014 to 0.0081 (overall=0.0048), as well as the gene diversity, from 0.533 to 1.000 (overall=0.945; Table 11). Sequences of the R2- dataset show 93 polymorphic sites and are, on average, slightly less variable: the overall sequence variability is equal to 0.0042 (varying from 0.0011 to 0.0116) and the overall gene diversity is 0.869 (ranging from 0.533 to 1.00; Table 11). It is to be noted that the 04-F27 individual shows an R2- mean p-distance value of 0.0116, significantly higher than the population average (0.0042). Grubb's test for outliers resulted significant for that value ($P<0.05$); therefore, it has been excluded from subsequent analyses.

For the downstream region, 133 28S sequences 810 bp long were obtained: 63 for the R2+ dataset and 70 for the R2- one. In the former alignment, 83 sites were found variable, whereas in the latter, 95 were polymorphic. Also in this region, sequence and gene diversities vary widely: 0.0015–0.0067 and 0.800–1.000, respectively, for the R2+ dataset; 0.0012–0.0086 and 0.786–1.000, respectively, for the R2- one (Table 11).

Overall, R2+ and R2- values for both parameters are almost the same: sequence variability is equal to 0.0041 for both dataset; gene diversity results 0.929 and 0.943, respectively. Student *t*-tests performed on both sequence and gene diversity measures did not show any significant comparison between R2+ and R2- datasets.

R2 presence does not seem to interfere with the 28S homogenization process. Moreover, tests conducted between individuals carrying the complete element and those without it again did not show any significant comparison, for both R2+ and R2- datasets and for both upstream and downstream regions. Tajima's D test performed on single R2+ and R2- alignments, as well as for

pooled datasets, rejects the neutrality hypothesis in the majority of trials, especially for the upstream region; moreover, all values, whether significant or not, are negative (Table 11).

Table 11: Mean sequence variability (p-D), gene diversity (H), and Tajima's D (per individual and overall) of inserted (R2+) and uninserted (R2-) 28S rDNA units.

	p-D		H		*Tajima's* D		
	R2+	*R2–*	*R2+*	*R2–*	*R2+*	*R2–*	*All*
Upstream							
04-M2	0.0014	0.0022	0.533	0.778	−1.741*	−1.873*	−2.316**
04-M3	0.0039	0.0011	0.972	0.533	−1.889*	−1.667	−2.394**
04-M4[a]	0.0027	0.0027	0.972	0.833	−1.822*	1.822*	−2.386**
04-M6[a]	0.0081	0.0054	0.867	0.893	−2.009**	−1.807*	−2.502**
04-F24	0.0036	0.0037	0.972	1.000	−1.795*	−1.756*	−2.338**
04-F25	0.0035	0.0036	1.000	0.917	−1.795*	−1.876*	−2.415**
04-F27	0.0070	0.0116	0.978	1.000	−1.913*	−1.518	−2.275**
04-F32[a]	0.0035	0.0031	0.972	0.917	−1.667	−1.436	−2.275**
Overall	*0.0048*	*0.0042*	*0.945*	*0.869*	*−2.873****	*−2.859****	*−2.932****
Downstream							
04-M2	0.0027	0.0041	0.917	0.978	−1.729	−1.710	−2.252**
04-M3	0.0029	0.0051	0.972	1.000	−1.632	−1.937*	−2.420***
04-M4[a]	0.0015	0.0012	0.800	0.786	−0.447	−1.534	−1.810*
04-M6[a]	0.0067	0.0086	1.000	1.000	−1.211	−1.495*	−2.178**
04-F24	0.0059	0.0037	1.000	0.972	−1.967**	−1.629	−2.444***
04-F25	0.0022	0.0040	0.722	0.833	−1.797*	−1.745*	−2.368**
04-F27	0.0029	0.0042	0.933	1.000	−1.390	−1.642	−2.130**
04-F32[a]	0.0045	0.0027	1.000	1.000	−1.444	−1.421	−2.187**
Overall	*0.0041*	*0.0041*	*0.929*	*0.943*	*−2.832****	*−2.866****	*−2.927****

[a]Samples with no full-length R2 element. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

3.2 The R2 element in Italian and Austrian T. cancriformis

3.2.1 Structure

For both Italian and Austrian *T. cancriformis* only the 3' portion of the element was determined because it isn't present the complete element. A Southern blot was done to assess the presence of the complete element, even if not visible on the electrophoretic gel, and the result was negative.

The R2sequence obtained from Italian population (R2Tc Italy) is 1841 bp long and exhibits an A+T content equal to 56%; the 3' end revealed a poly-A tail of 3 nucleotides. It contains a 1560 bp long ORF, located between nucleotide 1 and nucleotide 1560, coding for 520 amino acids comprising a reverse transcriptase domain and an endonuclease domain (Figure 22).

The partial sequence of the R2 element from the Austrian population (R2Tc Austria) is a consensus sequence of 6 clones 1004 bp long and exhibits an A+T content equal to 54%; the 3' end revealed a poly-A tail of 3 nucleotides. The ORF is 699 bp long, located between nucleotide 1 and nucleotide 699, coding for 233 amino acids. This sequence doesn't comprise the reverse transcriptase domain, but it's present the endonuclease domain (Figure 22).

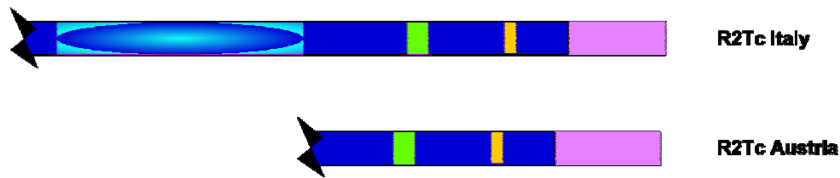

Figure 22: Schematic representation of the R2 sequences obtained for Italian and Austrian samples of *T. cancriformis*. Symbols and colours as in Fig. 20; black triangles indicate the end of the sequenced portion.

3.3 The R2 elements in L. lubbocki

3.3.1 Structure

Two elements were retrieved from the genome of *L. lubbocki*. They were indicated as R2Ll1 and R2Ll2.

To construct the complete sequence of R2Ll1, 6 clones for the 3' terminus and 5 clones containing the insertion site at the 5' terminus were considered. The sequence of the full-length R2 element is a consensus of the sequenced fragments obtained by primer walking. It is 4487 bp long and exhibits an A+T content equal to 58%. The sequencing of the R2 5' end showed the deletion of the two Gs of the 28S gene in all analyzed clones; no poly-A tail occurs at the 3' end. The R2Ll1 sequence contains a 2870 bp long ORF, located between nucleotide 1057 and nucleotide 3927, coding for 956 amino acids. This ORF shows a zinc-finger motif, the c-myb motif, the reverse transcriptase domain and the endonuclease domain (Figure 23). This ORF is interrupted by a stop codon at base 2362.

The complete sequence of R2Ll2 is based on 4 clones containing the whole insertion site at the 5' terminus and 13 clones for the 3' terminus. The sequence of the full-length element is a consensus of the sequenced fragments obtained by primer walking. It is 3649 bp long and exhibits an A+T content equal to 56%. Sequences of the 5' end showed an intact insertion site (TTAAGG) and the sequencing of the 3' end revealed a poly-A tail of 3 nucleotides. It wasn't possible to determine the starting codon of the functional ORF, and thereby the length of the 5' UTR. Moreover, the 5' portion of the element is characterized by frequent mutations that determine stop codons. Even if it isn't possible to recognize a complete ORF, R2Ll2 sequence contains a 1538 bp long ORF, located between nucleotide 1716 and nucleotide 3255, coding for 512 amino acids. The ORF presents the typical reverse transcriptase and endonuclease domains (Figure 23).

The comparison between R2Ll1 and R2Ll2 nucleotidic sequences (aligning the two sequences with BLAST) revealed that base 876 of R2Ll1 corresponds to base 22 of R2Ll2; of 3628 alignable nucleotides, 73% are identities (2683) and 1% are gaps (47). As far as the amino acidic sequence is

concerned, of 487 aligned positions 387 (79,5%) are conserved, 97 (20%) are variable sites and 3 are gaps.

Figure 23: Schematic representation of *L. lubbocki* R2 elements. Symbols as in Fig. 20. The black triangle indicates the stop codon.

3.3.2 Truncation analysis

The results of Southern blots on R2 truncation patterns are summarized in Table 12 and 13.

Table 12: Truncation variants profiles of R2Ll1. Numbers on the left represent the length in bp of each variant; individuals are indicated on the top (see Materials and Methods); **X** represents presence of a truncated variant, empty spaces represent its absence.

	RM M2	T36 M1	R11 F3	R11 M1	PEP F4	PEP F6	PEP M13	T36 F1	PEP F3	PEP M1
127	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
352	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
458	**X**		**X**	**X**		**X**	**X**	**X**	**X**	**X**
606	**X**		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Table 13: Truncation variants profiles of R2Ll2. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	RM M2	T36 M1	R11 F3	R11 M1	PEP F4	PEP F6	PEP M13	T36 F1	PEP F3	PEP M1
183	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
290	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
491	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
704	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
867		**X**	**X**		**X**	**X**		**X**	**X**	
943	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**
1025	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
1130	**X**	**X**	**X**	**X**	**X**				**X**	**X**
1270	**X**		**X**	**X**	**X**		**X**	**X**	**X**	**X**
1350			**X**		**X**	**X**			**X**	**X**
1546	**X**	**X**	**X**	**X**	**X**		**X**		**X**	
1627	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
2288	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Owing to T36M1 individual (lacking the complet element and truncated sequences) there are not fixed (e.g. present in all individuals) variants of R2Ll1. Excluding T36M1 sample, 3 out of 4 variants are fixed. On the whole variability level appears low.

R2Ll2 shows a total of 13 variants and each individuals bears from 9 to 13 variants; seven variants (183, 290, 491, 704, 1025, 1627, 2288) are fixed. Individuals R11 F3, PEP F4 and PEP F3 presents all truncated variants. Also in this case, the variability level is very low.

In both cases there aren't private variants, that is variants present in only one individual.

3.4 The R2 element in L. arcticus

3.4.1 Structure

The complete sequence was obtained assembling six clones for the 5' terminus and 3 clones for the 3' terminus. The sequence of the full-length R2 element (R2La) is a consensus of all sequenced fragments obtained by primer walking. It is 3557 bp long and exhibits an A+T content equal to 56,4%. The sequencing of the R2 5' end showed a deletion of the two Gs of the insertion site; the analysis of the 3' end revealed a poly-A tail of three nucleotides. The R2 sequence contains a 2862 bp long ORF, located between nucleotide 177 and nucleotide 3039, coding for 1013 amino acids. The ORF has a single zinc-finger motif, the c-myb domain, the retrotranscriptase domain and the endonuclease domain (Figure 24).

Figure 24: Schematic representation of *L. arcticus* R2 element. Symbols as in Fig. 20.

3.4.2 Truncation analysis

The results of Southern blots on R2 truncation patterns are summarized in Table 14 and 15.

Table 14: Truncation variants profiles of R2La of the Holtavörðuheiði population. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; X represents presence of a variant, empty spaces represent absence.

	Hol1	Hol10	Hol11	Hol12	Hol13	Hol14	Hol15	Hol16	Hol17	Hol18	Hol19
218	**X**	**X**									
411	**X**	**X**				**X**				**X**	**X**
729		**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		
1044							**X**	**X**	**X**		
1247					**X**						
1480											**X**

In the population of Holtavörðuheiði R2La is present with 6 variants, ranging from 1 to 3 per individual. There isn’t any fixed variant, while two private variants are scored: 1247 for individual Hol13 and 1480 for individual Hol 19.

Table 15: Truncation variants profiles of R2La of the Arnavatnsheiði population. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	Arn1	Arn2	Arn3	Arn4	Arn5	Arn6	Arn7	Arn8	Arn9	Arn10	Arn11	Arn12	Arn13	Arn14	Arn15	Arn16	Arn17
218				X													
411	X	X	X	X			X			X	X	X	X	X			X
729	X	X	X	X	X	X	X	X	X	X	X	X	X	X			X
1044								X	X	X		X	X			X	
1247										X				X			

In the population of Arnavatnsheiði the same 5 variants scored for Holtavörðuheiði (but the 1480 one) are retrieved. Each individual has from 1 to 4 truncated variants. Also here there isn’t any fixed variant and only a private one is present: locus 218 for individual Arn4. It is further to be noted the occurrence, as for R2Ll1 element, of an individual (Arn15) that lacks the complete element and truncated variants.

On the whole the two populations share 5 variants on 6; there isn’t any fixed variant and the most widespread (locus 729) is shared by 23 individuals.

3.5 The R2 elements in L. couesii

3.5.1 Structure

The sequencing of clones obtained from amplicons (R2IF1 > 28SB-R_L) of *L. couesii* revealed the presence of three R2 elements. They were named R2Lc1, R2Lc2 and R2Lc3. Specific features of each element are given in Table 16.

Table 16: Specific features of the 3’ portion of each element.

Element	**R2Lc1**	**R2Lc2**	**R2Lc3**
Total length (bp)	2106	2172	2147
A+T content (%)	57,3	54,3	57,4
3' end	AAA	AAA	AAA
ORF (bp)	1-1602	1-1599	1-1602
ORF (aa)	534	533	534

All the ORFs present the reverse transcriptase domain, a CCHC motif and the endonuclease domain.

Figure 25: Schematic representation of *L. couesii* R2 elements. Symbols as in Fig. 20; black arrowheads indicates work in progress.

The comparison of nucleotidic sequences of the three elements ORFs evidenced 806 conserved sites on 1602 sites. Of the 775 variable sites, the 27,4% of substitutions is in first position, the 18,6% in second position and the 54% is in third position. The pairwise comparison (Table 17) evidenced that R2Lc1 and R2Lc3 are more similar to each other with respect to R2Lc2.

The comparison of amino acidic sequences of the three elements evidenced the presence of 294 conserved sites on 534 sites. The pairwise comparison (Table 17) confirms data from nucleotidic sequences.

Table 17: Pairwise comparison of the nucleotidic (above the diagonal) and amino acidic (below the diagonal) sequences of the three elements. Numbers indicates the conserved sites, while in brackets are percentages of p-D.

	R2Lc1	R2Lc2	R2Lc3
R2Lc1	-	308 (41)	463 (11,2)
R2Lc2	677 (43,1)	-	307 (41)
R2Lc3	288 (18,3)	673 (42,7)	-

3.6 The R2 element in B. rossius

3.6.1 Structure

The complete sequence of *B. rossius* R2 element is a consensus sequence obtained from three clones for the 3' portion and six clones for the 5' portion. It's 3515 bp long and has an A+T content equal to 47%; the insertion site is characterized by the deletion of the two Gs, while the 3' end has a poly-A tail of about 23 nucleotides. The ORF is 3165 bp long (from base 228 to base 3392) and encodes for 1054 amino acids (Fig. 25).

Figure 26: Schematic representation of *B. rossius* R2 element. Symbols as in Fig. 20.

3.6.2 Truncation analysis in field collected specimens

Truncation variants profiles of Anzio, Patti, Curcuraci, Bertinoro, Massa San Nicola and Castanea delle Furie populations are given in Tables 18, 19, 20 and 21 (pp. 59-62).

43 truncated variants were scored in total, ranging from 2 to 25 per individual. There aren't fixed variants shared by all populations, but each population has its own fixed variants (Table 22).

Table 22: The total number of variants scored per population, the number of them that is shared by all individuals in the population (fixed), the number of them that is present only in a given population (private), the number of fixed variants that are also private of a given population and the number of individuals screened (N) are given.

	Total	Fixed	Private	Fixed/Private	N
Anzio	40	1	12	/	21
Patti	18	3	1	1	18
Curcuraci	14	1	/	/	19
Bertinoro	13	11	/	/	2
Massa San Nicola	15	12	/	/	4
Castanea delle Furie	15	13	/	/	4

From data in Table 22, it clearly emerges that gonochoric populations present a higher number of truncated variants, with Anzio being the far more variable population. Further, only gonochoric samples show fixed and private variants, even if with very different frequency.

The parthenogenetic population of Curcuraci, despite being constituted only by females, has a low number of fixed variants. On the contrary, the parthenogenetic populations of Bertinoro, Massa San Nicola and Castanea delle Furie have the higher number of fixed variants, but the fact that a very low number of individuals was analysed must be taken into account.

AMOVA test results indicated that the 58% of the variation is among populations, while the 42% is within populations; the Phi_{ST} value is 0,58.

PCA analyses (Fig. 26) revealed 3 well defined populations (Curcuraci, Patti and Anzio) and 3 populations (Bertinoro, Massa San Nicola, Castanea delle Furie) that are separated, although very similar to each other.

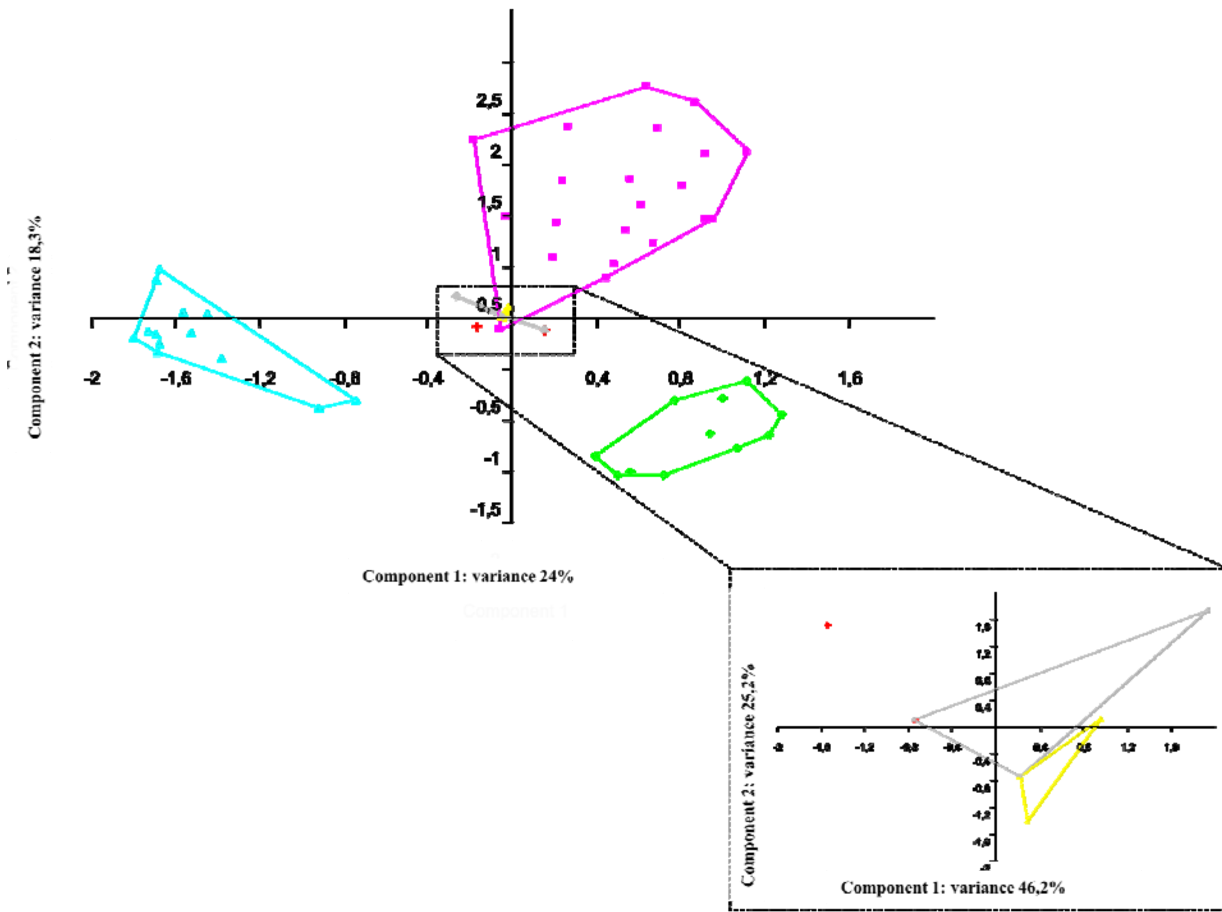

Figure 27: PCA analyses based on the presence/absence matrix of truncated variants. Next to each component the percentage value of variance is indicated. In the dotted box the particular of Bertinoro, Massa San Nicola and Castanea delle Furie populations is showed. Green: Curcuraci; light blue: Patti; red: Bertinoro; grey: Massa San Nicola; yellow: Castanea delle Furie; pink: Anzio.

3.6.3 Truncation analysis in laboratory crosses specimens

The screening of truncated variants in offspring (repeated also on parents) was conducted using three couples of primers: 28S-F2 > R2R2RIN, 28S-F2 > R2R2RIN2 and 28S-F2 > R2R2RIN6, because I saw that these were the most informative couples. Nevertheless there are discrepancies between variant profiles of the parents, in particular in the second analyses, conducted with the offspring, because some bands weren’t detected. For the descendant variants analyses, I took into account only bands that, for the parents, were scored in both screenings or that were present at least in one descendant. These preliminary analyses must be completed with the screening of the individuals with primer couples 28S-F2 > R2R2RIN4 and 28S-F2 > R2R2RIN5.

In table 23 (p. 63) the truncated variants profiles of parthenogenetic females 2 and 3 from Curcuraci and the respective offspring are given. In tables 24 and 25 (pp. 64 and 65) the truncated variants

profiles of female 5 from Curcuraci and male 15 from Anzio and of Female 6 from Curcuraci and male 9 from Anzio and the respective offspring deriving from their cross are given.

Individual BrCu♀2 has 8 variants and its offspring maintain from 4 to 7 of them; 4 variants are fixed. Individual BrCu♀3 has 5 variants and its offspring maintain from 3 to 4 of them; 3 variants are fixed. In both cases, no new variants are detected.

Female 5 from Curcuraci and male 15 from Anzio have respectively 4 and 8 variants; males deriving from their cross maintain from 2 to 8 of them. Two variants, present in both parents, are fixed in the offspring. Female 6 from Curcuraci and male 9 from Anzio have respectively 4 and 3 variants; males deriving from their cross maintain from 1 to 3 of them. One variant, present in both parents, are fixed in the offspring. A new variant (598) is scored in male 7.

3.7 The R2 element in B. grandii grandii

3.7.1 Structure

The R2 element of *B. grandii grandii* (R2Bgg) is a consensus sequence obtained from 6 clones for the 3' portion and 7 clones for the 5' portion; it's 3513 bp long and has an A+T composition of 46,6%. The 3' end has a poly-A tail of 7 nucleotides. The insertion site is deleted, since the sequence TTAAGGT is missing.

The R2Bgg sequence contains two ORFs (Figure 26) that are partially overlapping; the first ORF is in frame +2. It is 1698 bp long (from base 1700 to 3397) and encode for 565 amino acids. The second ORF is in frame +3 and is 1545 bp long (from base 171 to 1715) and encodes for 514 amino acids; it has a single zinc-finger motif and the c-myb motif. BLAST analyses of these two ORFs revealed that both have high levels of identity with R2 elements present in GenBank, but since the RT domain is included in the first one, I used this for phylogenetic analyses.

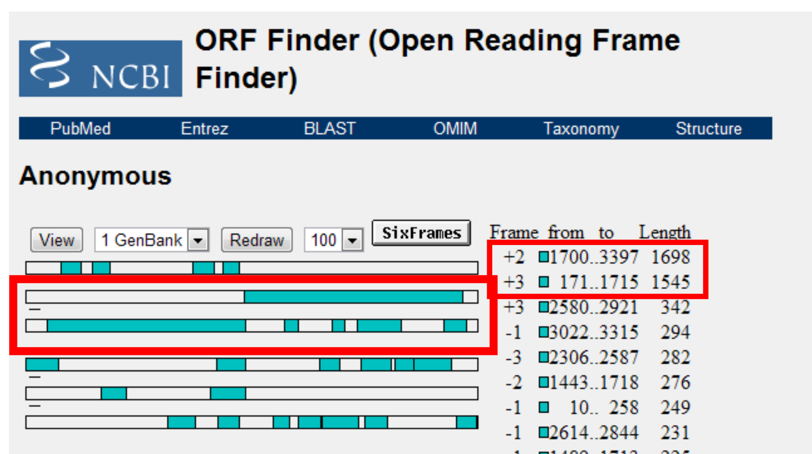

Figure 28: ORF finder output for the R2Bgg sequence. Results described in text are indicated with red boxes.

3.8 Summary

Complete elements that have also complete ORFs are R2Tc Spain, R2La and R2Br, while R2Ll1, R2Ll2 and R2Bgg are complete elements that encodes incomplete ORFs. R2Tc Italy and R2Tc Austria are incomplete elements that have also incomplete ORFs. The total length of the three elements of *L. couesii* -R2Lc1, R2Lc2 and R2Lc3- is still unknown because the work on them is in progress.

The structural features and the schematic representations of the R2 elements described above are summarized in Table 26 and Figure 29.

Table 26: Structural features of the R2 elements sequenced.

Element	Total length (bp)	A+T content (%)	3' end	ORF (bp)	ORF (aa)
R2Tc Spain	3583	53	AAA	177-3272	1031
R2Tc Italy	1841	56	AAA	1-1560	520
R2Tc Austria	1004	54	AAA	1-699	233
R2Ll1	4487	58	/	1915-4109	730
R2Ll2	3649	56	AAA	1716-3255	512
R2La	3557	56,4	AAA	177-3039	1013
R2Lc1	2106	57,3	AAA	1-1602	534
R2Lc2	2172	54,3	AAA	1-1599	533
R2Lc3	2147	57,4	AAA	1-1602	534
R2Br	3515	47	(A)23	228-3392	1054
R2Bgg	3513	46,6	(A)7	1700-3397	565
R2Bgg				171-1715	514

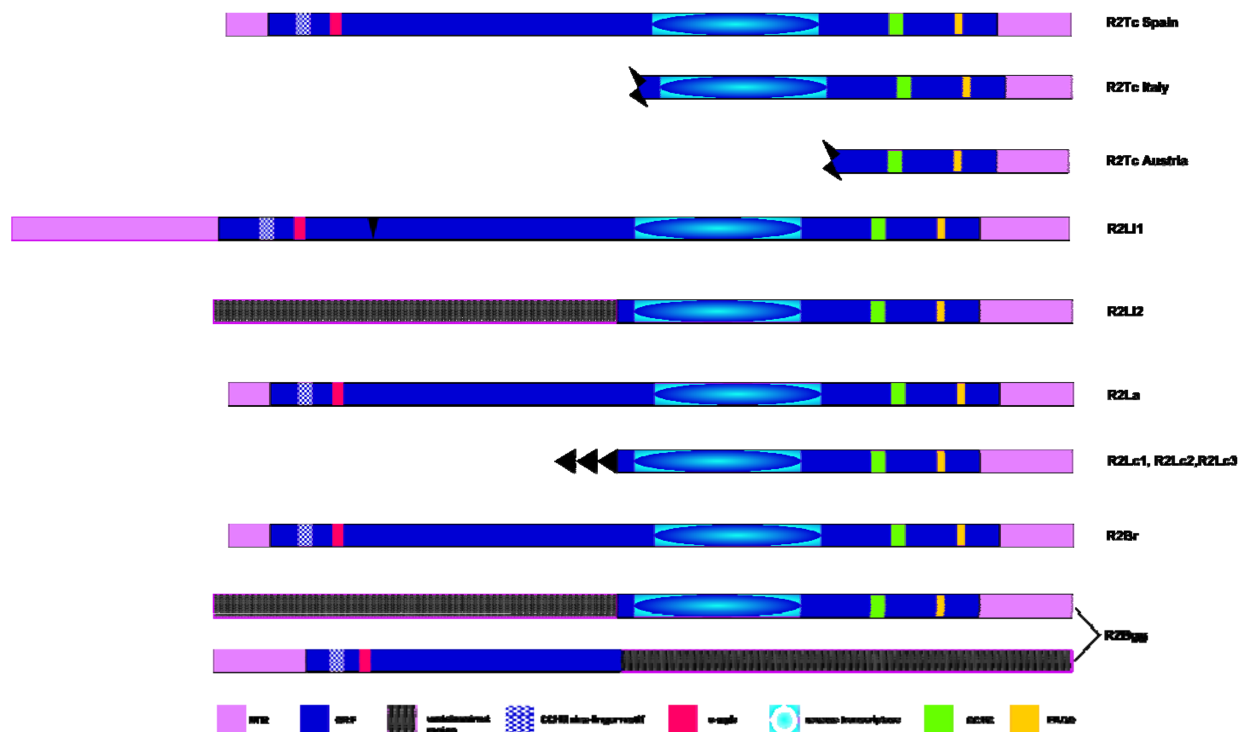

Figure 29: Schematic representation of the R2 elements characterized.

3.9 Phylogenetic analyses

The phylogenetic analyses were performed on amino acid sequences of the ORFs; in particular only the portion -of each element- that starts from the RT domain till the end of the ORF was used, because it's demonstrated that's the most phylogenetic-informative part (Malik et al., 1999; Burke et al., 1999; Eickbush, 2002; Kojima and Fujiwara, 2005).

The starting alignment is the one presented in Mingazzini et al. (2010), to which R2 sequences from termites (published in Ghesini et al., 2011) and the R2 sequences that I've characterized were added.

The same terminal branching pattern is observed for all phylogenetic estimation methods, even if with different node support values. In the Neighbor Joining and Bayesian dendrograms (Figure 30), clades A, B, C, and D and subclades found by Mingazzini et al. (2010) can be recognized with some differences possibly occurring owing to the addition of new sequences. The main variation is given by the presence of a new subclade, that I have named D7, composed by branchiopods R2 elements, except *T. longicaudatus*: in fact R2Tl still lies in the A1 subclade (Kojima and Fujiwara, 2005; Mingazzini et al., 2010). The subclade D7 is very well supported by Bayesian posterior probability (0,99) and its branches are well resolved, because for all internal nodes the support value is maximum. R2 elements of *T. cancriformis* from Austria and Italy (R2Tc Austria and R2Tc Italy) cluster together and are the basal couple of the entire subclade. R2Tc from Spain isn't

comprised in the *Triops* cluster, but it is basal to all *Lepidurus* R2 elements. The second R2 element of *L. lubbocki* (R2Ll2) is basal to the cluster formed by the first R2 element of *L. lubbocki* (R2Ll1) and the second element of *L. couesii* (R2Lc2). Finally, the first element of *L. couesii* (R2Lc1) is basal to the couple composed by the element of *L. arcticus* (R2La) and the third element of *L. couesii* (R2Lc3).

R2 elements of stick insects (R2Br and R2Bgg) cluster together in all analyses with maximum support values and are comprised in the D5 subclade. It is to be noted the position of the element of the fourth tick species (R2Amo, *A. monolakensis*), lying in clade B instead of clade A, even if with very low support values (see Discussion).

Figure 30: R2 phylogeny inferred by Neighbor Joining and Bayesian methods. Number right to each node indicates the bootstrap value, as a percentage of 1000 replicates, while numbers in italic left to each node indicates the posterior probability (values below 0,95 aren't indicated since not significative). Letters indicate clades and subclades (as described in the text). Light blue box highlights branchiopods R2 elements here characterized, while green box highlights stick insects R2 elements.

Table 18: Truncation variants profiles of *B. rossius* females from Anzio. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	An F1	An F2	An F3	An F4	An F5	An F18	An F19	An F20	An F21	An F22	An M6	An M7	An M8	An M9	An M10	An M11	An M12	An M13	An M14	An M15	An M16
113		X		X	X	X		X	X	X	X	X	X	X		X					X
164	X			X	X	X			X												
222	X			X	X	X			X	X	X				X				X	X	X
291	X								X												
364	X			X	X	X		X	X	X	X				X		X		X	X	X
439	X			X	X	X				X	X				X	X	X		X	X	X
490	X		X	X	X	X	X	X			X		X		X	X	X		X	X	X
536	X		X	X	X	X		X	X	X	X										
598											X		X		X	X					
658	X								X		X				X	X					
715			X	X		X		X								X					
767	X			X		X			X	X	X				X		X			X	X
839																					
870																					
880	X	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
958	X		X	X		X		X		X					X		X	X	X	X	
1058	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1147	X				X	X			X	X	X		X		X	X	X		X		
1218			X	X	X		X	X					X			X		X	X	X	X
1265	X	X	X	X					X	X	X	X	X	X	X	X	X	X	X	X	X
1317	X		X				X			X		X	X	X	X	X	X		X		X
1383			X	X		X		X													
1416	X				X					X	X				X				X		X
1479	X	X	X	X	X		X			X	X	X	X	X	X	X	X	X	X	X	X
1574	X		X			X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
1655											X		X		X				X		
1711										X				X	X				X		
1761											X		X	X	X	X	X	X	X		
1829				X	X	X	X		X	X		X	X	X	X	X	X	X	X	X	
1907					X				X	X	X			X	X	X	X	X	X		X
2025	X			X		X	X	X			X	X	X	X	X	X	X	X	X	X	
2127												X	X	X	X	X	X	X	X		
2121	X																				
2197			X				X		X	X	X					X				X	
2217																					
2270															X		X				
2322					X			X		X	X	X	X	X	X		X		X		
2407				X		X													X		
2476	X		X				X												X		
2525										X	X		X	X			X				
2582								X		X											
2619							X														
2742					X																

Table 19: Truncation variants profiles of *B. rossius* individuals from Patti. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	Pa F1	Pa F2	Pa F3	Pa F4	Pa F12	Pa F13	Pa F14	Pa F15	Pa F16	Pa F17	Pa M5	Pa M6	Pa M7	Pa M8	Pa M9	Pa M10	Pa M11	Pa M18
113	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
164																		
222																		
291	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
364																		
439														X				
490																		
536																		
598																		
658																		
715																		
767		X	X	X	X	X						X	X		X			
839	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
870	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
880														X			X	
958							X	X	X	X								X
1058	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
1147	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	
1218	X	X	X	X	X	X								X	X	X	X	
1265							X	X										
1317																		
1383																		
1416																		
1479			X	X	X		X	X	X	X	X			X	X	X	X	
1574																		
1655	X	X	X	X	X	X	X	X	X	X	X			X	X	X	X	
1711																		
1761																		
1829																		
1907																		
2025	X	X	X	X	X	X	X	X	X	X	X			X	X			
2127	X	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X
2121																		
2197				X														
2217	X	X	X	X	X	X	X	X			X			X		X		
2270																		
2322																		
2407							X	X										
2476																		
2525																		
2582																		
2619																		
2742																		

Table 20: Truncation variants profiles of *B. rossius* females from Curcuraci. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	Cu F1	Cu F2	Cu F3	Cu F4	Cu F5	Cu F6	Cu F7	Cu F8	Cu F9	Cu F10	Cu F11	Cu F12	Cu F13	Cu F14	Cu F15	Cu F16	Cu F17	Cu F18	Cu F19
113																			
164	X	X	X	X															
222	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
291																			
364	X	X	X	X		X	X					X							
439																			
490																			
536																			
598																			
658																			
715																			
767	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
839																			
870																			
880																			
958																			
1058			X	X	X	X	X												
1147																			
1218			X	X	X														
1265																			
1317																			
1383																			
1416	X	X	X	X						X	X	X							
1479	X	X	X	X	X	X	X	X	X										
1574	X	X	X	X	X	X	X	X	X	X	X	X							
1655																			
1711																			
1761	X	X	X	X	X	X	X	X	X	X	X	X		X				X	X
1829																			
1907																			
2025																			
2127																			
2121																			
2197	X	X	X	X	X	X	X	X	X	X	X	X					X		
2217																			
2270													X	X	X	X	X		
2322																			
2407																			
2476													X	X	X	X	X		
2525																			
2582																			
2619																			
2742													X	X	X	X	X		

Table 21: Truncation variants profiles of *B. rossius* individuals from Bertinoro, Castanea delle Furie and Massa San Nicola.. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	Bert1	Bert2	CdF1	CdF2	CdF3	CdF4	MSN1	MSN2	MSN3	MSN4
113			X	X	X	X	X			X
164										
222	X	X	X	X	X	X	X	X	X	X
291										
364		X	X	X	X	X	X	X	X	X
439										
490										
536										
598										
658				X						
715										
767	X	X	X	X	X	X	X	X	X	X
839										
870										X
880										
958										X
1058	X	X	X	X	X	X	X	X	X	X
1147	X									
1218	X	X	X	X	X	X	X	X	X	X
1265										
1317										
1383						X				X
1416										
1479										
1574	X	X	X	X	X	X	X	X	X	X
1655	X	X	X	X	X	X	X	X	X	X
1711										
1761	X	X	X	X	X	X	X	X	X	X
1829	X	X	X	X	X	X	X	X	X	X
1907										
2025	X	X	X	X	X	X	X	X	X	X
2127	X	X	X	X	X	X	X	X	X	X
2121										
2197										
2217	X	X	X	X	X	X	X	X	X	X
2270										
2322										
2407										
2476										
2525										
2582										
2619										
2742										

Table 23: Truncation variants profiles of *B. rossius* females (BrCu♀2 and BrCu♀3) and their respective offspring. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	BrCu♀2	F1	F2	F3	F4	F5	F164	F169	F170	F172	F173	BrCu♀3	F1	F2	F4	F5	F6	F12*	F13*	F14*	F15*	F16*
113																						
164	X	X	X	X	X	X	X	X				X	X	X	X	X	X	X	X	X	X	X
222	X	X	X	X	X	X	X	X														
291																						
364	X	X	X	X	X	X	X	X	X	X	X											
439																						
490																						
536																						
598																						
658																						
715																						
767	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
839																						
870																						
880																						
958																						
1058												X	X	X	X	X	X	X	X	X	X	X
1147																						
1218												X				X			X			X
1265																						
1317																						
1383																						
1416	X	X	X	X	X	X	X	X	X	X	X											
1479	X	X	X	X	X	X	X	X	X	X	X											
1574												X			X		X					
1655																						
1711																						
1761	X				X																	
1829																						
1907																						
2025																						
2127																						
2121																						
2197	X					X																
2217																						
2270																						
2322																						
2407																						
2476																						
2525																						
2582																						
2619																						
2742																						

Table 24: Truncation variants profiles of *B. rossius* Curcuraci ♀5, Anzio ♂15 and males offspring deriving from the cross. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	Cu ♀5	An ♂15	M1	M4	M8	M9	M10	M133	M134	M143	M144	M176
113												
164												
222	X	X			X	X	X	X		X	X	
291												
364		X					X		X			X
439		X		X			X		X		X	X
490		X		X			X		X		X	X
536												
598												
658												
715												
767	X	X	X	X	X	X	X	X	X	X	X	X
839												
870												
880		X		X			X		X		X	X
958												
1058												
1147												
1218	X	X	X	X	X	X	X	X	X	X	X	X
1265												
1317												
1383												
1416												
1479												
1574												
1655												
1711												
1761												
1829												
1907												
2025												
2127												
2121												
2197	X	X					X					
2217												
2270												
2322												
2407												
2476												
2525												
2582												
2619												
2742												

Table 25: Truncation variants profiles of *B. rossius* Curcuraci ♀6, Anzio ♂9 and males offspring deriving from the cross. Numbers on the left represents the length in bp of each variant; individuals are indicated on the top; **X** represents presence of a variant, empty spaces represent absence.

	Cu ♀6	An ♂9	M16	M19	M22	M24	M25	M171	M172	M173	M174	M185
113												
164												
222	X						X	X			X	
291												
364												
439												
490												
536												
598									X			
658												
715												
767	X	X	X	X	X	X	X	X	X	X	X	X
839												
870												
880												
958												
1058												
1147												
1218												
1265												
1317		X				X			X			
1383												
1416												
1479												
1574	X	X	X									
1655												
1711												
1761												
1829												
1907												
2025												
2127												
2121												
2197	X			X	X		X	X		X	X	
2217												
2270												
2322												
2407												
2476												
2525												
2582												
2619												
2742												

4. Discussion

4.1 R2 elements structure

The R2 elements of *T. cancriformis*, *L. arcticus* and *B. rossius* that I have sequenced are complete elements that well match general features of this kind of transposable elements. So, a single ORF with a central RT domain is present, as in type II non-LTR retrotransposons (Eickbush and Jamburuthugoda, 2008); upstream the RT domain a single zinc-finger motif and the c-myb motif can be found, while downstream the RT the endonuclease domain is present (Burke et. al., 1999). It could be seen that the structural features of these three elements are well conserved. The R2La and R2Br 5' ends present the deletion of the two Gs of the 28S insertion site, while R2Tc Spain has more extended deletions; their 3' ends show the typical poly-A tail. These features are consistent with previous studies on *Drosophila spp*. (George et al., 1995; Burke et al., 1999; Perez-Gonzalez and Eickbush, 2001) and are determined by the target primed reversed transcription mechanism, in the phase of DNA cleavage and cDNA synthesis (Christensen et al., 2006).

For *L. lubbocki*, complete sequences were retrieved; their ORFs, though, is interrupted by stop codons in the 5' portion. For R2Ll1 it's possible to determine the length of the 5' UTR, that is 6 times longer than that of other R2 elements, and to recognize the zinc-finger and c-myb motifs. On the contrary, for R2Ll2 it isn't possible to determine the length of the 5'-UTR and the putative start codon, because of the frequent stop codons. However, the 3' portion of the ORF containing the RT domain is intact in both elements. The incomplete ORF of R2Ll2 doesn't seem to codify for a a functional product, the element being therefore non-functional. On the other hand, R2Ll1 has a complete ORF, even if this is interrupted. Considering that my sequence is only a sample of all complete elements present in the *L. lubbocki* genome, it is possible to suggest the existence of a complete functional R2Ll1 without a stop codon that interrupts the ORF.

Even considering that the majority of elements reported in the literature are only partially sequenced, this is the first case that a complete element with an incomplete ORF (R2Ll2) is found (Kojima and Fujiwara, 2005). Stage and Eickbush (2009, 2010) reported the occurrence of incomplete R2 elements in *D. mojavensis* and in *Nasonia spp*., but they explain the finding with a lack of sequence coverage. In my case, it appears a demonstration that inactive elements can remain in the genome and accumulate mutations.

The R2 sequence of the *B. grandii grandii* element is the most difficult to interpret, because it has two ORFs partially overlapping, in different frames and both well matches the general features of R2 elements. The fact that these two ORFs aren't separated and that the endonuclease domain is downstream the RT domain, led me to the conclusion that this isn't a R1 element (see chapter 4.4

for specific features). The only example of an unusual R2 element comes from *N. vitripennis*: in fact R2NvA is thought to be a fusion between R1 and a R2 elements (Burke et al., 1993; Stage and Eickbush, 2010), however this is not my case. The most likely explanation for R2Bgg is the recombination between two R2 elements, but obviously further analyses are needed.

The situation of R2 from Italian and Austrian *T. cancriformis* is different, because there's no evidence of the presence of complete elements, so the obtained sequences correspond to putatively inactive elements.

Finally, R2 elements from *L. couesii* are only partially sequenced and their ORFs have typical features of R2 elements; their complete sequencing is in progress.

4.2 28S variation analysis in T. cancriformis

The presence of R2 (either truncated or not) within a 28S gene can influence the ribosomal sequence homogeneity in two ways: (i) a large insertion (kilobases long) may interfere with recombination, preventing the pairing of inserted with uninserted rDNA repeats; (ii) once R2 is inserted within a 28S sequence, the ribosomal gene becomes unfunctional and may freely accumulate mutations. In both instances, R2 insertion should hinder concerted evolution, basically avoiding the homogenization process. It could, therefore, be expected that inserted sequences would be more variable than the uninserted ones. In *T. cancriformis*, the presence of R2 does not impact on 28S sequence and gene diversity. Indeed, sequence divergence comparisons between R2+ and R2- 28S genes are not significant, showing very close variability values. This is in line with data on *Drosophila*, in which inserted and uninserted 28S are identical: this has been explained through the rapid elimination of new insertions (Eickbush and Eickbush, 2007). Differently, 28S rRNA genes carrying the Pokey element in *D. pulex* accumulate more mutations than those without the insertion: the authors explain this contrasting pattern as the results of the long persistence (or even the spreading) of some Pokey insertions preventing ribosomal unit recombination and homogenization (Glass et al., 2008). The R2 turnover suggested by data on *T. cancriformis* do not support this dynamic, as the newly transposed variants appear to be rapidly eliminated. As argued by Glass et al. (2008), recently generated 28S inserted copies would be indistinguishable from those that never experienced the element insertion and, because of their quick elimination, they do not significantly alter the rDNA homogenization level. The generally low variability observed in this analysis suggests a quite efficient process of sequence conservation and the hypothesis of neutrality has been rejected in several instances. Interestingly, the majority of significant Tajima's Ds can be observed in the region upstream of the insertion site, whereas in only few instances this has been shown in the downstream region. The downstream region here characterized is homologous to that described

by Glass et al. (2008) in *D. pulex* in which the same, non-significant, values have been observed: this may indicate that this region undergoes neutral evolution. Generally speaking, all Tajima's D values observed here are negative, evidencing an excess of low frequency polymorphisms: this can be either the results of purifying selection (that can be expected, of course) or caused by a recent expansion of new 28S variants. Measures of gene diversity are consistent with the latter scenario as the higher values obtained are expected when there are several alleles none of which reaching very high frequency. This well reconciles with the R2 turnover observed in *T. cancriformis*: as for any retrotransposition event large rDNA units deletions occur (Zhang et al., 2008), a compensatory replacement of new 28S variant is necessary for proper functionality. Multiple cycles of rDNA unit gains and losses would boost their turnover, leading to a quite homogeneous array (as rDNA units are very recently duplicated), but at the same time let spread several single point mutations throughout the array. It would be interesting to investigate if, in the absence of R2, the same pattern of sequence and gene diversity can be achieved. The three individuals without the complete (active) R2 element show no difference in comparison with those having the full-length retrotransposon; however, as they were sampled from a random mating population, it is unlikely that subsequent generations will lack the active element. The R2/rDNA 'interplay' can be interpreted on a Red Queen (Salathé et al., 2008) view: new niches for the "parasite" R2, more efficient homogenization of rDNA units to contain its spread, and so on. Does this dynamics bring further advantage to the host? The turnover of rDNA units within the array, in which the retrotransposition occurs, may lead to a greater variance in the proportion of functional/defective 28S rRNA genes between individuals. In an evolutionary perspective, this would result in more opportunities for natural selection to operate on the host.

4.3 R2 phylogeny

Earlier studies outlined an important aspect of R2 evolution: with only few exceptions, the R2 and host phylogenies do not overlap. Two hypotheses have been put forward to explain this pattern: in the first, vertical inheritance of the element can be followed by lineage extinction or diversification in certain groups; the second hypothesis assumes the horizontal transfer of R2 between species. In a recent survey, the former has been shown as the most likely explanation (Kojima and Fujiwara, 2005), so that the incongruence between host and R2 phylogeny can be explained, almost totally, by high rates of diversification of the element and not by horizontal transfer between species. The matter is very intriguing, because deep nodes of R2 phylogeny are consistent with structural features (number of zinc-finger motifs at the N-terminus; Kojima and Fujiwara, 2005) and, if we assume a vertical inheritance followed by diversification, some still-unknown factor should underlie

this consistency. Obviously, the two hypotheses (vertical vs horizontal transmission) may not be mutually exclusive. All the elements that I have characterized fall in the D clade (Kojima and Fujiwara, 2005), characterized by the occurrence of a single zinc-finger motif. This is confirmed by sequence data for R2Tc Spain, R2Ll1, R2La, R2Br and the first ORF of R2Bgg, but it's not verifiable for the other elements presented in this thesis with only partial sequences. Branchiopods R2 elements form a monophyletic cluster that does not comprise the *T. longicaudatus* element, which lies within the group of elements with three zinc-finger motifs (Clade A). Actually, this is not currently verifiable because the only available *T. longicaudatus* R2 sequence is not complete. It's possible that R2Tl diversified itself from the common ancestor of *Triops* R2 lineage, due to the low fidelity of the reverse transcriptase (see chapter 5.3). It remains to be clarified how an element can gain (or loose) zinc-finger motifs independently and/or randomly with respect to host phylogeny.

The level of identity between *L. lubbocki* elements and among *L. couesii* elements is comparable to that found among elements of the turtle species *Mauremys reevesii* (R2CrA, R2CrB1 and R2CrB2). Even elements from different congeneric species (e.g. *Drosophila spp*. and *Reticulitermes spp*.) show percentages of similarity higher than *Lepidurus spp.* elements. In the phylogenetic analyses, R2Ll1 and R2Ll2 don't cluster together. The same applies to R2Lc1, R2Lc2 and R2Lc3 elements. On the whole they can be considered as different elements.

Stick insects R2 elements (R2Br and R2Bgg) cluster together in all elaboration, with maximum support values. Their position in clade D is corroborated by the presence of one zinc-finger motif at their 5' end. It would be interesting to extend R2 sequencing to other stick insect species, in order to observe the phylogenetic behaviour of the element.

It is to be noted that R2 from the tick *A. monolakensis* (R2Amo) isn't comprised in the clade that groups R2s from other tick species (clade D6) and presents an ambiguous position. In previous elaborations (Bunikis and Barbour, 2005; Mingazzini et al., 2010) it was part of clade A, while in present analyses it is comprised in clade B. Its collocation is currently not verifiable, because the available sequence is incomplete and it's not possible to count the number of zinc-fingers at 5' end. furthermore, in the paper where it was first presented, its position was ambiguous too: in fact, in phylogenetic trees published by Bunikis and Barbour (2005), R2Amo falls in clade A, but without support values. Authors explained this situation with two hypotheses: the first is the presence of paralogous elements in soft ticks (with respect to hard ticks, see clade D6), while the second one is the acquisition of a new R2 lineage. Of course, complete sequencing is needed to clarify this point.

Owing to the low fidelity of its reverse transcriptase, R2 sequences are subject to a high diversification leading to the occurrence of multiple lineages within the same species and/or elimination of some lineages owing to competition for the limited number of insertion sites (Pérez-

Gonzalez and Eickbush, 2001; Eickbush and Jamburuthugoda, 2008). Moreover, R2s are ancient components of the animal genome, their presence dating back at least to the splitting of cnidarians and bilaterians (Kojima et al., 2006). The antiquity and the evolutionary dynamics of R2 may explain, therefore, the lack of correlation between its phylogeny and that of the host species. Nevertheless, it's possible that, beside "real" incongruences, artefact ones are present. For example, R2Tc, in Mingazzini et al. paper (2010), was comprised in subclade D5; then, extending the data set with sequences from other branchiopods, it's seen that they form a monophyletic clade, always except R2Tl. So R2Tl position seems to be a real incongruence, while the R2Tc first placement was an artefact incongruence.

Therefore, I think that beyond the horizontal transfer/diversification/extinction of elements, a cause of R2 phylogeny incongruences can be the exiguity of the data set or the false phylogenetic signal possibly given by mutated, and possibly inactive, elements.

4.4 Truncation analysis

My studies on R2 truncation variants are the first on populations different from *Drosophila spp*. Organisms and their populations are chosen because, beside gonochoric reproduction, they show non-canonical reproduction modalities such as hermaphroditism and parthenogenesis. Until today there aren't data about the activity of R2 in low-recombining or non- recombining genomes, so my data represent a contribution to the knowledge of the dynamics of retroelements depending on the reproduction modality. Results here presented and discussed represent the starting point and analyses will be extended to a wider number of samples.

R2 truncation analyses have been so far conducted only on laboratory stocks of *D. melanogaster* and *D. simulans*. In the former, the variants distribution has been found, to some extent, well conserved, with ancestral-truncated variants being shared by individuals both within and between isofemale lines. However, some lines of *D. simulans* show decidedly higher R2 activity, producing less conserved truncation profiles (Pérez- Gonzalez and Eickbush, 2001, 2002; Pérez-Gonzalez et al., 2003; Zhang and Eickbush, 2005). A recent survey on natural populations of *D. simulans* showed a high turnover rate, each individual carrying a specific collection of R2 truncations (Zhou and Eickbush, 2009). The high incidence of R2 insertions in *D. simulans* is correlated with a high rate of variant elimination and a low number of inserted 28S, explainable as due to its retrotransposition creating large deletions in adjacent rDNA units, thus eliminating a number of R2 variants (Zhang et al., 2008).

The dynamics of R2 in gonochoric populations of *T. cancriformis* is in line with these observations as (i) individuals from the same and/or different samples show very different truncation profiles, (ii)

there are not ancestral variants shared by all individuals, and (iii) the percentage of rDNA units with insertions is very low (0.5–5%). However, a peculiarity related to the R2 elimination occurs: six tadpole shrimps show truncated variants, but not the complete element. Therefore, a complete R2 is lacking in their genomes.

The case of R2 element in parthenogenetic (Italian) and hermaphroditic (Austrian) populations of *T. cancriformis* can represent an intermediate step toward the extinction of the element, given that nor any complete element, neither any truncated variant are scored. The particular situation observed in the parthenogenetic population can be due to its unisexual reproductive strategy and is partially in agreement with the studies of Gladyshev and Arkhipova (2009) on the Bdelloid rotifer *Adineta vaga*. Authors reported that the majority of R9 retroelement copies are not functional and in general retroelements are present in the rotifer genome in a so low number, that it isn't possible to detect them with PCR screenings. Therefore I cannot exclude that a complete element exist, but its elimination, and the elimination of its truncated variants, is significantly faster than its activity.

The apparent absence of an active R2 in the Austrian hermafroditic population is a difficult topic to discuss, because my expectation was that R2 can have the same activity of Spanish population. Zhang and Eickbush (2005) found that different *D. simulans* isofemale lines can bear active or inactive elements at random, and suggested the presence of cellular systems that can detect the expression of any sequence present in multiple copies. These cellular systems are substantially transcriptional and post-transcriptional silencing mechanisms (see chapter 4.3), but it remains an unresolved point why in some individuals these mechanisms are activated and in others are inactive.

On the whole, *T. cancriformis* R2 elements seems to be active in gonochoric populations and inactive in parthenogenetic and hermaphroditic ones; moreover, it is possible that in the latter populations R2s are at the end of their "lifecycle", as described by Schaack et al. (2010). Authors said that TEs lifecycle is a sort of "birth and death" process in which elements "die" in genomes when all their copies are eliminated or inactivated (by cellular mechanisms or by mutation accumulation that also prevent the annealing of specific primers).

As for *T. cancriformis*, truncation analyses concerning R2 elements from parthenogenetic *L. lubbocki* are conducted at the population level. Individuals analysed constitute a subsample of populations collected in different years, and represent the starting point for future wider analyses. My data don't enable me, at present, to determine the ancestral condition of both elements: R2Ll1 and R2Ll2, in fact, show a "frozen" truncation pattern, that is variants shared by most of the individuals and absence of private variants. Variants absence is distributed by chance among individuals and doesn't have a clear temporal progress, I mean it isn't clear if individuals lack variants because of the rDNA turnover (real deletion) or because they descent from a parent that

lacks the variant. The pattern of R2 variants seems to indicate that R2Ll1 and R2Ll2 are likely to be inactive; this is corroborated also by the fact that R2Ll1 and R2Ll2 have non-functional ORFs (see chapter 8.1). These data are comparable with those of Ghesini et al. (2011), who found no evidence of retrotransposition activity in termites, and with those of Pérez-Gonzalez and Eickbush (2002, 2003) who reported similar pattern for R2 in *D. melanogaster* and concluded that ancestral variants in their isofemale lines must be present since the isolation of each line and that variants elimination rate, by means of concerted evolution, is threefold the retrotransposition rate. On the whole, data on R2 structure and truncation analyses support the hypothesis made on "birth and death" of retrotransposon families and let to the conclusion that these elements seems to be inactive and on the way of the extinction in non-recombining genomes. Finally, the absence of R2Ll1 in an individual deserves explanation; however it could be related to the "birth and death" theory of retrotransposon families and can represent a step toward the extinction of the element.
The situation of R2 in *L. arcticus* is slightly different, since a single functional element is scored that's potentially active. Even so, individual truncation profiles in the two egg-bearing individuals populations reflects low variability as for *L. lubbocki*. The low number of variants scored and the absence of the complete element in an individual is in line with data discussed above for parthenogenetic populations and confirm the trend of "frozen" retroelements activity in non-recombining genomes.
The truncation analysis conducted on *B. rossius* R2 from field collected populations put in light some unusual aspects. It clearly emerges that the element is active in the gonochoric population of Anzio, that has the 93% of the scored variants and only one is fixed. The gonochoric population of Patti has 3 fixed variants and the others are shared by most of the individuals. Unexpectedly, the truncation pattern of the parthenogenetic population of Curcuraci is similar to that of Patti and is different from those of the other parthenogenetic populations, since it has a single fixed variant, as it's for Anzio gonochoric population, even if the majority of the variants are shared among many individuals. Parthenogenetic populations of Bertinoro, Massa San Nicola and Castanea delle Furie have lower truncation variability, but with a higher number of ancestral variants fixed in all individuals. A possible explanation of the difference existing between Curcuraci females and other parthenogenetic populations is that Curcuraci population was gonochoric in the recent past (Scali V., oral communication) and now is in a transition phase to obligate parthenogenesis, while females from the other parthenogenetic populations are respectively siblings and variants are fixed simply because they descent from the same females.
PCA analysis output indicate that there are 3 well defined/separated populations (Anzio, Patti and Curcuraci) and that put in light the absence of shared truncated variants. It's expected under the

assumption of concerted evolution of the rDNA locus: the elimination rate of R2Br variants is higher than retrotransposition events and molecular drive that act on the rDNA locus fix different variants in different populations. The high level of fixed variants in parthenogenetic populations from Bertinoro, Massa San Nicola and Castanea delle Furie and the absence of private ones well fit observations made on bdelloid rotifers by Arkhipova (2005) and on *D. pulex* by Valizadeh and Crease (2008) and Schaack (2010). In these papers they report a minor presence/inactivation of TE in parthenogentic organisms. Nevertheless, it would be necessary to expand the sample.

Analyses on laboratory born specimens of *B. rossius* indicate that offspring, either derived from parthenogenetic females or from crosses between Curcuraci females and Anzio males, generally doesn't share all variants with the respective parents, but instead have deletions of variants. Only one individual (M172) from the cross between female 6 from Curcuraci and male 9 from Anzio presents a new insertion. Generally speaking, my expectation was to score a higher number of new truncated variants, due to the potential activity of the element. or to the sum of variants present in the parents; on the other hand the deletions evidence a rapid and efficient turnover of rDNA units. Truncated variants analyses on first generation offspring can give an indication on R2 behaviours in outcrosses, however further investigations on next generations offspring are needed to detect a clear pattern of element activity.

A general picture of the situation suggests a complex interaction between transposition-mediated deletions, genomic turnover and silencing mechanisms that act in different ways depending also on the reproductive strategy of the organism considered. In *D. simulans*, inserted 28S are hypothesized to be eliminated by the transposition of active R2 elements, whereas genomic turnover mechanisms tend to replace deleted rDNA units with new ones for the maintenance of the ribosomal locus functionality. This, however, also creates new niches for the R2 element, which can remain active (Eickbush and Eickbush, 2007; Zhang et al., 2008). On the other hand, as a consequence of transposition-mediated deletions, the loss of R2 variants, either complete or truncated, might be dramatic: the element copy number can be reduced to very few copies (for example, a single 28S rDNA with an insertion). Moreover, genomic turnover mechanisms acting on rDNA locus might eliminate all 28S carrying insertions. Therefore, a process such as transposition-mediated deletion, together with unequal DNA exchanges acting on the few 28S units carrying the complete R2 elements, can explain why they can be lost (e.g. in the 15% of the gonochoric *T. cancriformis,* 10% of parthenogenetic *L. lubbocki* and 3,6% of egg-bearing *L. arcticus* assayed). Once the full-length copy is deleted, new insertions cannot occur and the remaining truncations would be progressively eliminated by subsequent rounds of genomic turnover mechanisms. The loss of the R2 element from a genome is the first step toward the extinction in a given population/species, possibly leading

to unclear phylogenetic patterns; it is, therefore, essential to understand how this mechanism proceeds.
The elimination of R2 through the interplay between transposition-mediated deletions, genomic turnover and silencing mechanisms might give a clue to the process, but how the absence of R2 can be maintained in a population is still a further, open question. In a gonochoric population, as the Spanish tadpole shrimp samples are, outcrosses between individuals without the complete element and individuals carrying a functional R2 will very likely result in an offspring with active (complete) elements. This would mean that the extinction of an R2 lineage in a given population is an unlikely event. However it could happen, as it has been already shown (Jakubczak et al., 1991) and as my results corroborate.

5. Conclusions

The dynamic of R2 is influenced by many mechanisms that are intermingled and together drive the genome/element evolution.
My data are concordant with previous studies but also contribute to clarify some R2 evolutionary aspects, and even if further analyses are needed, some main points deserve to be considered:
- R2 is a multicopy element inserting in a repeated sequence: the rDNA gene 28S. My data on inserted/uniserted 28S variability confirm that ribosomal repeats follow a pattern of concerted evolution and undergo purifying selection (Perèz-Gonzalez and Eickbush, 2002);
- the presence of multiple lineages of R2 in the same species (*L. lubbocki*, *L. couesii*) suggests that the life cycle of this element follows a "birth and death" evolutionary pattern;
- the vertical inheritance of the element is the most likely mechanism of transmission, but events of horizontal transmission can not be ruled out (e.g. R2 from *T. longicaudatus*);
- the Muller's ratchet theory states that low and non-recombining genomes are likely to be extinct on the long term due to the accumulation of deleterious mutations. If I consider R2 as a deleterious mutation, I assume that, in theory, hermaphroditic and parthenogenetic populations must either don't have R2s at all or have R2s whom activity is out of control. What I have found seems to contrast the Muller's ratchet hypothesis, because hermaphroditic and parthenogenetic populations actually have R2s. These R2s are found to be incomplete (hermaphroditic and parthenogenetic *T. cancriformis*), unfunctional (R2s from parthenogenetic *L. lubbocki*) or seemingly "frozen" (egg bearing individuals of *L. arcticus* and parthenogenetic populations of *B. rossius*). It's clear that mechanisms that limit R2 spread act in low- and non-recombining genomes.

References

Aksoy S., Williams S., Chang S. and Richards F. F., 1990. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. Nucleic Acids Res 18: 785–792.

Arkhipova I. R., 2005. Mobile genetic elements and sexual reproduction. Cytogenet. Genome Res. 110 (1-4): 372-82.

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

Brtek J. and Thiéry A., 1995. The geographic distribution of the European branchiopods (Anostraca, Notostraca, Spinicaudata, Levicaudata). Hydrobiologia 298: 263–80.

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 (5): 465-74.

Burke W. D., Calalang 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., Muller F., and Eickbush T. H., 1995. R4, a non- LTR retrotransposon specific to the large subunit rRNA genes of nematodes. Nucleic Acids Res. 23: 4628– 34.

Burke W. D., Singh D., Eickbush T. H., 2003. R5 retrotransposons insert into a family of infrequently transcribed 28S rRNA genes of planaria. Mol. Biol. Evol. 20 (8): 1260-70.

Burke W. D., Malik H. S., Jones J. P. and Eickbush T. H., 1999. The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol. Biol. Evol. 16: 502–11.

Burke W. D., Malik H. S., Rich S. M. and Eickbush T. H., 2002. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. Mol. Biol. Evol. 19 (5): 619-30.

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: 163–85.

Càceres C. E. and Soluk D. A., 2002. Blowing in the wind: A field test of overland dispersal and colonisation by aquatic invertebrates. Oecologia 131: 402–8.

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. U.S.A. 103: 17602–17607.

Dawkins R., 1976. The Selfish Gene. New York City: Oxford University Press.

Dolgin E. S. and Charlesworth B., 2006. The fate of transposable elements in asexual populations. Genetics. 174 (2): 817-27.

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

Dover G. A., 1982. Molecular drive: a cohesive mode of species evolution. Nature 299: 111-117.

Dover G. A., 2002. Molecular drive. Trends Genet. 18: 587-589.

Doyle J. J. and Doyle J. L., 1987. A rapid DNA isolation method for small quantities of fresh tissues. Phytochemical Bulletin 19: 11–15.

Eickbush T.H., 1994. Origin and evolutionary relationships of retroelements. In: The Evolutionary Biology of Viruses (pp. 121–157). Morse, S.S. (Ed.), Raven Press, New York,.

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

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: 221–34.

Eickbush T.H. and Malik H.S., 2002. Evolution of retrotransposons. In: Mobile DNA II (pp. 1111–44). Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), American Society of Microbiology Press, Washington DC.

Eickbush T. H., Burke W. D., Eickbush D. G. and Lathe III W. C., 1997. Evolution of R1 and R2 in the rDNA units of the genus Drosophila. Genetica 100: 49–61.

Figuerola J., Green A. J. and Michot T. C., 2005. Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. Am. Nat. 165: 274–280.

Fisher D. C., 1990. Rate of evolution in living fossils. In: Paleobiology (pp 152-159). Briggs DEG, Krauther PS (Eds), Blackwell Scientific, London.

Fryer G., 1988. Studies on the functional morphology and biology of the Notostraca (Crustacea: Branchiopoda). Philos T Roy Soc B 321: 27–124.

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): 184-191.

George J. A., Burke W. D. and Eickbush T. H., 1995. Analysis of the 5' junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. Genetics 142: 853-63.

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. Jan 23. [Epub ahead of print] DOI 10.1007/s00239-011-9430-y.

Gladyshev E.A. and Arkhipova I.R., 2007. Telomere-associated endonuclease deficient Penelope-like retroelements in diverse eukaryotes. Proc. Natl. Acad. Sci. U.S.A. 104: 9352–57.

Gladyshev E.A. and Arkhipova I.R., 2009. Rotifer rDNA-specific R9 retrotransposable elements generate an exceptionally target site duplication upon insertion. Gene 448: 145-50.

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.

Glass S. K., Moszczynska A. and Crease T. J., 2008. The effect of transposon Pokey insertions on sequence variation in the 28S rRNA gene of Daphnia pulex. Genome 51: 988–1000.

Gorokhova E., Dowling T. E., Weider L. J., Crease, T. J. and Elser J. J., 2002. Functional and ecological significance of rDNA intergenic spacer variation in a clonal organism under divergent selection for production rate. Proc. R. Soc. Lond B 269: 2373– 2379.

Hammer Ø., Harper D. A. T. and Ryan P. D., 2001. PAST: Paleontological Statistics software package for education and data analysis. Paleontologia Electronica 4 (1): 9 pp.

Hessen D. O., Rueness E. K. and Stabell M., 2004. Circumpolar analysis of morphological and genetic diversity in the Notostracan *Lepidurus arcticus*. Hydrobiologia 519:73–84.

Hickey D. A., 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 101: 519-31.

Hillis D. M. and Dixon M. T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quart. Rev. Biol. 66: 411- 453.

Huelsenbeck J. P., 2000. Mrbayes: Bayesian Inference of Phylogeny. Version 1.1. Available via http://brahms.biology.rochester.edu/software.html.

Jacubczack J. L., Burke W. D. and Eickbush T. H., 1991. Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc. Natl. Acad. Sci. USA 88: 3295- 99.

Kaplan N. L. and Brookfield J. F. Y., 1983. Transposable elements in Mendelian populations. III. Statistical results. Genetics 104: 485–495.

Kazazian H. H. Jr, 2004. Mobile elements: drivers of genome evolution. Science. 12; 303(5664): 1626-32.

Keeling P. J. and Palmer J. D., 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9 (8): 605-18.

Kidwell M. G. and Lisch D. R., 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55: 1-24.

King J. L. and Hanner R., 1998. Cryptic species in a 'living fossil' lineage: taxonomic and phylogenetic relationships within the genus *Lepidurus* (Crustacea: Notostraca) in North America. Mol. Phylog. Evol. 10: 23–36.

Kojima K. K. and Fujiwara H., 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol. 20(3): 351-61.

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

Kojima K. K., Kuma K., Toh H. and Fujiwara H., 2006. Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. Mol. Biol. Evol. 23: 1984–93.

Kordis D. and Gubensek F., 1999. Horizontal transfer of non-LTR retrotransposons in vertebrates. Genetica.;107(1-3):121-8.

Korn M. and Hundsdoerfer A. K., 2006. Evidence for cryptic species in the tadpole shrimp *Triops granarius* (Lucas, 1864) (Crustacea: Notostraca). Zootaxa 1257: 57–68.

Korn M., Marrone F., Pérez-Bote J. L., Machado M., Cristo M. and Cancela da Fonseca L., 2006. Sister species within the *Triops cancriformis* lineage (Crustacea, Notostraca). Zool. Scr. 35: 301–22.

Lampe D. J., Witherspoon D. J., Soto-Adames F.N. and Robertson H.M., 2003. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol. Biol. Evol. 20 (4): 554-62.

Le Rouzic A. and Capy P., 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 169 (2): 1033-43.

Lee Y. C. and Langley C. H., 2010. Transposable elements in natural populations of *Drosophila melanogaster*. Philos. Trans. R Soc. Lond. B Biol Sci. 27; 365 (1544): 1219-28.

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.

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 (8): 88.

Longhurst A. R., 1955. A review of the Notostraca. Bulletin of the British Museum of Natural History, D, 3: 1–57.

Luchetti A., Scanabissi F. and Mantovani B., 2006. Molecular characterization of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis* (Crustacea, Branchipoda, Notostraca). Genome 49: 888-93.

Luchetti A., Scanabissi F. and Mantovani B., 2007. Evolution of LEP150 sub-repeat array within the ribosomal IGS of the clam shrimp *Leptestheria dahalacensis* (Crustacea Branchiopoda Conchostraca). Gene 1; 400 (1-2): 174-80.

Lynch J. E., 1966. *Lepidurus lemmoni* Holmes: a redescription with notes on variation and distribution. Transaction of the American Microscopy Society 85: 181–92.

Lynch J. E., 1970. *Lepidurus couesii* Packard (Notostraca) redescribed with a discussion of specific characters in the genus. Crustaceana 23: 43–9.

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

Mantovani B. and Scali V., 1993. Genetic structure and phyletic relationships of eastern Mediterranean *Bacillus atticus* Brunner (Insecta, Phasmatodea). Biochem. Genet. 31: 343–62.

Mantovani B., Cesari M. and Scanabissi F., 2004. Molecular taxonomy and phylogeny of the "living fossil" lineages *Triops* and *Lepidurus* (Branchiopoda: Notostraca). Zoologica Scripta 33: 367-74.

Mantovani B., Cesari M. and Scanabissi F., 2009. Molecular taxonomy and phylogeny of Italian *Lepidurus* taxa (Branchiopoda: Notostraca). It. J. Zool. 76 (4): 358- 65.

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., Scali V. and Tinti F., 1991. Allozyme analysis and phyletic relationships of two new stick-insects from northwest Sicily: *Bacillus grandii benazzii* and *B. rossius–grandii benazzii* (Insecta, Phasmatodea). J. Evol. Biol. 4: 279-90.

Martin S. L. and Bushman F. D., 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol. Cell. Biol. 21 (2): 467-75.

Martirosyan I. A., Korchagin V. I., Tokarskaya O. N., Darevsky I. S. and Ryskov A. P., 2006. Finding of *Bov-B* LINE Retroelement in Parthenogenetic and Bisexual Lizard Species of the Genus *Darevskia* (Lacertidae). Russian Journal of Genetics, 42 (7): 790–94.

Mingazzini V., Luchetti A. and Mantovani B., 2010. R2 dynamics in *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution. Heredity, 2010 Epub ahead of print | doi:10.1038/hdy.2010.86.

Morales-Hojas R., Post R. J., Wilson M. D. and Cheke R. A., 2002. Completion of the sequence of the nuclear ribosomal DNA subunit of *Simulium sanctipauli*, with description of the 18S, 28S and the IGS. Medical and Veterinary Entomology 16: 386-94.

Murugan G., Maeda-Martìnez A. M., Obregòn-Barboza H. and Hernàndez-Saavedra N. Y., 2002. Molecular characterization of the tadpole shrimp *Triops* (Branchiopoda: Notostraca) from the Baja California Peninsula, México: new insights on species diversity and phylogeny of the genus. Hydrobiologia 486: 101–13.

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

Okamoto H. and Hirochika H., 2001.Silencing of transposable elements in plants. Trends Plant Sci. 6 (11): 527-34.

Orgel L. E. and Crick F. H., 1980. Selfish DNA: the ultimate parasite. Nature 284: 1517–23.

Pérez-Gonzalez C. E. and Eickbush T. H., 2001. Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans*. Genetics 158: 1557–67.

Pérez-Gonzalez C. E. and Eickbush T. H., 2002. Rates of R1 and R2 retrotransposition and elimination from the rDNA locus of *Drosophila melanogaster*. Genetics 162: 799–811.

Perez-Gonzalez C. E., Burke W. D. and Eickbush T. H., 2003. R1 and R2 retrotransposition and elimination from the rDNA loci of the X and Y chromosomes of *Drosophila melanogaster*. Genetics 165: 675–85.

Rasband W. S., 1997–2007. ImageJ. U.S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/.

Rogers D. C., 2001. Revision of the Nearctic Lepidurus (Notostraca). J. Crust. Biol. 21: 991–1006.

Ronquist F. and Huelsenbeck J. P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–74.

Rozen S. and Skaletsky H. J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Kravetz S, Misener S, editors. Bioinformatics methods and protocols: Methods in molecular biology. Totowa, NJ: Humana Press. pp. 365–86.

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.

Sambrook E., Fritsch F. and Maniatis T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press: Cold Spring Harbor, NY.

Sánchez-Gracia A., Maside X. and Charlesworth B., 2005. High rate of horizontal transfer of transposable elements in Drosophila. Trends Genet. 21 (4): 200-3.

Sassaman C., Simovich M. A. and Fugate M., 1997. Reproductive isolation and genetic differentiation in North American species of *Triops* (Crustacea: Branchiopoda: Notostraca). Hydrobiologia 359: 125–47.

Scali V. and Marescalchi O., 1987. The evolution of the genus *Bacillus* (Phasmatodea, Bacillidae) from a karyological point of view. Pp. 541-549 in B. Baccetti, ed. Evolutionary biology of orthopteroid insects. Horwood, Chichester, U.K.

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.

Scanabissi F. and Mondini C., 2002. A survey of the reproductive biology in Italian branchiopods. Part A. The female gonad of *Limnadia lenticularis* (Linnaeus, 1761) (Spinicaudata) and *Lepidurus apus lubbocki* Brauer, 1873 (Notostraca). Hydrobiologia 486: 263– 72.

Scanabissi F. and Mondini C., 2002. A survey of the reproductive biology in Italian branchiopods. Part B. The male gonad of *Lepidurus apus lubbocki* Brauer, 1873 (Notostraca). Hydrobiologia 486: 273–8.

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.

Scanabissi F., Alfonso G., Bergamaschi S. and Mantovani B., 2006. Primo ritrovamento di Lepidurus couesii Packard, 1875 in Italia. Thalassia Salentina 29:113–24.

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

Schaack S., Pritham E. J., Wolf A. and Lynch M., 2010. DNA transposon dynamics in populations of *Daphnia pulex* with and without sex. Proc Biol Sci. 277 (1692): 2381-7.

Schlüter P. M. and Harris S. A., 2006. Analysis of multilocus fingerprinting data sets containing missing data. Mol. Ecol. Notes 6: 569-72.

Srivastava A. K. and Schlessinger D., 1991. Structure and organization of ribosomal DNA. Biochimie 73 (6): 631-8.

Stage D. E. and Eickbush T. H., 2007. Sequence variation within the rRNA gene loci of 12 *Drosophila* species. Genome Res. 17 (12): 1888-97.

Stage D. E. and Eickbush T. H., 2009. Origin of nascent lineages and the mechanisms used to prime second-strand synthesis in the R1 and R2 retrotransposons of *Drosophila*. Genome Biol. 10: R49.

Stage D. E. and Eickbush T. H., 2010. Maintenance of multiple lineages of R1 and R2 retrotransposable elements in the ribosomal RNA gene loci of *Nasonia*. Ins. Mol. Biol. 19 (1): 37-48.

Suno-Uchi N., Sasaki F., Chiba S. and Kawata M., 1997. Morphological stasis and phylogenetic relationships in tadpole shrimps, *Triops* (Crustacea: Notostraca). Biol. J. Linn. Soc. 61: 439–57.

Swofford DL., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates: Sunderland, Massachusetts.

Tamura K., Dudley J., Nei M. and Kumar S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596–99.

Valizadeh P. and Crease T. J., 2008. The association between breeding system and transposable element dynamics in *Daphnia pulex*. J Mol Evol.66 (6): 643-54.

Volff J. N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays. 28 (9): 913-22.

Weider L. J., Elser J. J., Crease T. J., Mateos M., Cotner J. B. and Markow T. A., 2005. The functional significance of ribosomal rDNA variation: impacts on the evolutionary ecology of organisms. Annu. Rev. Ecol. Evol. Syst. 36: 219-42.

Wessler S. R., 2006 a. Eukaryotic transposable elements: teaching old genomes new tricks. In: The implicit genome (pp. 138-162). Oxford University Press Inc.

Wessler S. R., 2006 b. Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. USA. 103 (47): 17600-1.

Wingstrand K. G., 1978. Comparative spermatology of the Crustacea Entomostraca. 1. Subclass Branchiopoda. Det Kongelige Danske Videnskarbernes Selskab Biologiske Skrifter 22: 1-66.

Xiong Y. and Eickbush T. H., 1988. Similarity of reverse transcriptase-like sequences of viruses, ransposable elements, and mitochondrial introns. Mol. Biol. Evol. 5 (6): 675-90.

Xiong Y. and Eickbush T. H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9 (10): 3353-62.

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: 229–37.

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.

Supplementary materials

Mingazzini V., Luchetti A. and Mantovani B., 2010. R2 dynamics in *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution. Heredity, 2010 Epub ahead of print | doi:10.1038/hdy.2010.86.

Heredity (2010), 1–9
© 2010 Macmillan Publishers Limited All rights reserved 0018-067X/10 $32.00
www.nature.com/hdy

ORIGINAL ARTICLE

R2 dynamics in *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca): turnover rate and 28S concerted evolution

V Mingazzini, A Luchetti and B Mantovani
Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Bologna, Italy

The R2 retrotransposon is here characterized in bisexual populations of the European crustacean *Triops cancriformis*. The isolated element matches well with the general aspects of the R2 family and it is highly differentiated from that of the congeneric North American *Triops longicaudatus*. The analysis of 5′ truncations indicates that R2 dynamics in *T. cancriformis* populations show a high turnover rate as observed in *Drosophila simulans*. For the first time in the literature, though, individuals harboring truncation variants, but lacking the complete element, are found. Present results suggest that transposition-mediated deletion mechanisms, possibly involving genomic turnover processes acting on rDNAs, can dramatically decrease the copy number or even delete R2 from the ribosomal locus. The presence of R2 does not seem to impact on the nucleotide variation of inserted 28S rDNA with respect to the uninserted genes. On the other hand, a low level of polymorphism characterizes rDNA units because new 28S variants continuously spread across the ribosomal array. Again, the interplay between transposition-mediated deletion and molecular drive may explain this pattern.
Heredity advance online publication, 14 July 2010; doi:10.1038/hdy.2010.86

Keywords: concerted evolution; non-LTR retrotransposon R2; *Triops cancriformis*; ribosomal DNA; turnover rate; transposition-mediated deletion

Introduction

Non-LTR retrotransposons are transposable elements that can be either randomly distributed or inserted at a specific locus. Sequence specificity of insertion is considered an ancient strategy used by transposable elements to survive in the host genome by limiting their ability to disrupt essential genes (Malik *et al*., 1999). Ribosomal DNA represents a niche well exploited by non-LTR retrotransposons, and eight rDNA-specific families have been so far identified, with six of them inserting into the 28S gene (R1, R2, R4, R5, R6, RT; Eickbush and Eickbush, 2007).

One of the most studied non-LTR retrotransposon families is R2. Four clades have been so far recognized on the basis of the element's phylogeny and of the N-terminal zinc-finger motifs of the translated protein product (Kojima and Fujiwara, 2005). In particular, the R2-A, -C, and -D clades have 3, 2, and 1 zinc-finger motifs, respectively, whereas the N-terminal structure of the R2-B clade is, as yet, undetermined.

R2 occurs in the four triploblastic phyla Platyhelminthes, Arthropoda, Echinodermata, and Chordata, but its presence also in the diploblastic phylum Cnidaria suggests its vertical inheritance since the cladogenesis of Radiata and Bilateria (Kojima *et al*., 2006). Yet, R2 phylogeny is quite inconsistent with that of the host (Burke *et al*., 1999; Kojima and Fujiwara, 2005), only some taxonomic 'subclades' (*sensu* Kojima and Fujiwara, 2005; for example the *Drosophila* sp. and fishes–turtles subclades) being evident in the trees. Therefore, the hypothesis of horizontal transfer of the element has been put forward. However, its absence in taxa closely related to species harboring R2 indicates that the extinction of this retrotransposon has occurred several times, at least during insect and vertebrate evolution (for example, in *Drosophila erecta*, *Drosophila orena*, *Fugu rubripes*, mouse, and human). On the other hand, some species such as *Popillia japonica* and *Ciona intestinalis* have multiple lineages of the element (Burke *et al*., 1993; Eickbush *et al*., 1997; Kojima and Fujiwara, 2004). On the whole, both extinction and diversification can be explained by R2 evolutionary dynamics, showing a rapid turnover with high rates of retrotransposition and elimination (Pérez-Gonzalez and Eickbush, 2001, 2002; Zhang and Eickbush, 2005).

R2 inserts through a *target primed reversed transcription* mechanism (Christensen *et al*., 2006), which allows also the insertion of 5′-truncated copies; these are produced when the synthesis of R2 first-strand DNA is aborted before reaching the 5′ end of the element. The study of truncation variants is a tool to examine element activity: this aspect was deeply analyzed in laboratory stocks of *Drosophila* spp. In *Drosophila simulans*, in particular, a high turnover rate, together with transposition-mediated deletions, is responsible of the elimination of earlier generated truncation variants (Zhang *et al*., 2008). Moreover,

Correspondence: Professor B Mantovani, Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Via Selmi 3, Bologna 40126, Italy.
E-mail: barbara.mantovani@unibo.it
Received 6 March 2009; revised 30 March 2010; accepted 25 May 2010

the selective pressure against non-functional rDNA units tends to eliminate the R2-inserted copies through the unequal DNA exchanges acting in the concerted evolution of the ribosomal locus (reviewed in Nei and Rooney, 2005; Eickbush and Eickbush, 2007). Concerted evolution explains the variability pattern observed for repeated sequence families (such as ribosomal genes): the observed sequence variability within an evolutionary unit (a species, a subspecies or a population) is significantly lower than between different evolutionary units of the same rank. Concerted evolution is achieved through molecular drive, a process comprising the intragenomic homogenization of variants, through turnover mechanisms such as unequal crossing-over, gene conversion and rolling circle replication, and variant fixation within a group of reproductively linked bisexual organisms (Dover, 1982, 2002).

No detailed data are so far available about the effects of R2 insertion on sequence variability of 28S genes: in *Drosophila*, R2-inserted and non-inserted 28S units appear identical (Eickbush and Eickbush, 2007). However, in the crustacean *Daphnia pulex*, the DNA-mediated element *Pokey* determines a higher variability of the inserted 28S sequences with respect to those lacking the element (Penton and Crease, 2004; Glass *et al*., 2008).

We here characterize the R2 element in the Euroasiatic notostracan *Triops cancriformis*. The order Notostraca pertains to the class Branchiopoda, a primitive group of the Arthropoda sub-phylum Crustacea, recently placed within the Pancrustacea clade, strongly associated with Hexapoda (Halanych, 2004; Mallat *et al*., 2004). *T. cancriformis* is a well-known example of the very few living fossils: from its morphological stasis, it cannot be distinguished from the Triassic taxon *T. cancriformis minor* (Fisher, 1990). *T. cancriformis* inhabits ephemeral ponds and rice fields and shows a consistent variability in sexual reproductive strategies, which range from bisexuality (either gonochoric or hermaphroditic) to unisexuality (parthenogenesis; Mantovani *et al*., 2004, 2008). Here, we characterize the R2 element and analyze its turnover/elimination rates together with the 28S rDNA unit variation in Spanish gonochoric populations of *T. cancriformis* (*sensu* Korn *et al*., 2006).

Materials and methods

R2 molecular characterization and phylogenetic analysis

Tadpole shrimps were collected in Espolla (Spain): the same pond was sampled twice in 2004 and in 2006. Forty individuals were analyzed, 20 for each sampling (Table 1). Genomic DNA was extracted from single alcohol-preserved individuals with a standard phenol-chloroform protocol.

Samples were first checked for the presence of the R2 element using the forward degenerate primers described in Kojima and Fujiwara (2005), coupled with a 28SB-R reverse primer (Table 2), located 178 bp downstream of the element's insertion site. Of the tested primer pairs, only R2IF1 > 28SB-R gave an amplification product of the expected size (~2000 bp); this was cloned and sequenced as described below. PCR amplifications were performed in a 50 μl reaction mixture using the *TaKaRa LA Taq™ with GC Buffer* kit (TAKARA BIO Inc., Shiga, Japan), following the manufacturer's instructions. Thermal cycling was

Table 1 List of *T. cancriformis* samples used in this study and summary of truncation variants

Year	*Sex*	*Alias*	*Complete R2*	*Truncations number*	*Total*	*Mean*
2004	Male 1	04 M1	+	28		
2004	Male 2	04 M2	+	20		
2004	Male 3	04 M3	+	13		
2004	Male 4	04 M4	−	11		
2004	Male 5	04 M5	+	17		
2004	Male 6	04 M6	−	6		
2004	Male 7	04 M7	+	13		
2004	Male 8	04 M8	+	16		
2004	Male 9	04 M9	+	5		
2004	Male 10	04 M10	+	4		
2004	Female 22	04 F22	+	22		
2004	Female 23	04 F23	+	17		
2004	Female 24	04 F24	+	16		
2004	Female 25	04 F25	+	19		
2004	Female 27	04 F27	+	25		
2004	Female 28	04 F28	+	17		
2004	Female 29	04 F29	+	17		
2004	Female 30	04 F30	+	18		
2004	Female 31	04 F31	+	20		
2004	Female 32	04 F32	−	14	318	15.9
2006	Male 1	06 M1	+	11		
2006	Male 2	06 M2	+	3		
2006	Male 3	06 M3	+	16		
2006	Male 4	06 M4	+	12		
2006	Male 5	06 M5	+	8		
2006	Male 6	06 M6	+	18		
2006	Male 7	06 M7	+	18		
2006	Male 8	06 M8	+	10		
2006	Male 9	06 M9	−	20		
2006	Male 10	06 M10	−	20		
2006	Female 11	06 F11	+	6		
2006	Female 12	06 F12	+	10		
2006	Female 13	06 F13	+	17		
2006	Female 14	06 F14	+	6		
2006	Female 15	06 F15	+	16		
2006	Female 16	06 F16	+	9		
2006	Female 17	06 F17	−	2		
2006	Female 18	06 F18	+	20		
2006	Female 19	06 F19	+	10		
2006	Female 20	06 F20	+	7	239	11.95
					557	13.93

+/− indicates the presence/absence of the complete element.

Table 2 List of primers prepared for this study by the authors

Primer name	*Sequence 5′ > 3′*
1618F	GAAAGGGAATCCGGTTCCCATTCC
28S-F2	GTCAAAGTGAAGAAATTCAACGAAG
28SB-R	CGTCTCCCACTTATGCTACACCTC
28S-OUT	TTCAGGTATAATCAGACGGACGTAG
DIN-coda	AAGTGGGAAGTGTTTTCAATGTACT
DIN	GGGTATTCAATTCTCGCATCTC
DIN3	AAGAGTCCTCAACAAAATTTTAAACCTACT
DIN4	TACAAAGAGCTCGTTAAAGATCAGC
DIN5	GGATAAGAGTAAGTGTTCTGTTTGTGG
RIN	GCAGGGAAAAAGAGGCATTAG
RIN2	GAACTCCAACTCTAAACAAGAGGTATCAG
RIN3	CTAGGTAGGAGTTAGTCAAGTCAAGCAG
RIN4	GATCTCTCAAATGAAGGAGTAGGTTTA
RIN5	CAATAATGTTGTCAAGTTTGTGTTCTA
RIN6	CACTGATAAATCAGCTTACCCAGTCT
18-5′	CCTGGTTGATCCTGCCAG

For present analyses also the following primers were used: 18i (Hillis and Dixon, 1991), R2IF1, R2IF2, R2IIF1, and R2IIF2 (Kojima and Fujiwara, 2005).

94 °C for 5′, 94 °C for 30″, 48 °C for 30″, and 70 °C for 10′ for 35 cycles; 15′ at 72 °C as a final extension. Amplified PCR products were gel purified and cloned into a pGEM-T Easy vector (Promega, Madison, WY, USA). Sequencing was performed at Macrogen Inc. (Seoul, Korea). The complete sequence of R2 was obtained through the primer walking method. A total of 48 clones were sequenced to analyze the 5′ junction of the element with the 28S gene. Sequences were edited and assembled using MEGA4 (Tamura *et al*., 2007).

Quantification of rDNA units and R2 copies within the *T. cancriformis* genome was performed through dot blot analysis. Genomic DNA was spotted onto positively charged nylon membranes (GE Healthcare Limited, Amersham Place, Little Chalfont, Buckingamshire, UK) in a series of dilutions (2000-15.6 ng); probe lanes had dilutions ranging from 5 to 0.04 ng for the 18S probe, and from 0.1 to 0.00078 ng for the R2 probe. To quantify the percentage of rDNA units, the blotted membrane was hybridized with a 400 bp long 18S probe obtained using primers 18-5′ > 18i (Table 2). To score the percentage of R2-inserted units, the filter was hybridized with a 1309 bp probe specific for R2 (primer pair DIN > RIN; Table 2). Hybridizations were performed under highly stringent conditions, with the final wash at 65 °C in 0.1 × SSC, 0.1% SDS. Probe labeling and blot detection were performed using the *DIG High Prime DNA Labeling and Detection Starter Kit I* and the *CDP-Star* (Roche Diagnostics GmbH, Mannheim, Germany). Images were analyzed using ImageJ (Rasband, 1997–2007).

The open reading frame was found using the ORF Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The phylogenetic analysis was performed on amino-acid sequences using the alignment of Kojima and Fujiwara (2005), to which all GenBank available R2-encoded proteins were added: *Blattella germanica* (Accession number: EF014490; Kagramanova *et al*., 2007), *Amblyomma americanum* (AY682792), *Boophilus microplus* (AY682793), *Ixodes scapularis* (AY682794), *Argas monolakensis* (AY682796) (Bunikis and Barbour, 2005), and *Nematostella vectensis* (Kojima *et al*., 2006). Amino-acid sequences were aligned with the MAFFT software online version (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/) using the G-INS-i algorithm with BLOSUM62 matrix. Neighbor Joining and Maximum Parsimony dendrograms were computed using PAUP* 4.0b10 (Swofford, 2003), with gaps treated as informative characters; bootstrap values were obtained after 1000 replicates. The Bayesian phylogenetic tree was constructed using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). Monte Carlo Markov chains ran for 2 million generations, with trees sampled every 100 generations; the first 332 trees were discarded as burn-in. In all analyses, the SLACS element (CAA34931) of *Trypanosoma brucei* (Aksoy *et al*., 1990) was used as outgroup.

Evaluation of R2 activity

R2 activity was analyzed through 5′ truncation patterns as described in Pérez-Gonzalez and Eickbush (2001). Truncated element copies were obtained by PCR amplification (see above) using the 28S-F2 primer, annealing 62 bp upstream of the element insertion site, coupled with various R2-specific primers: RIN2, RIN3, RIN4, and RIN5 (Table 2). These primers anneal 3020, 1899, 1251, and 699 bp downstream of the insertion site, respectively. Individuals were separated into four groups consisting of 10 specimens each: 2004 females, 2004 males, 2006 females, and 2006 males. All groups were screened with all primer pairs. A total of 10 μl of each amplification product were separated on a 2% agarose gel and Southern blotted onto a positively charged nylon membrane (GE Healthcare). For every primer pair, a specific probe was designed and used to hybridize with the related PCR product: DIN > RIN2 (963 bp); DIN3 > RIN3 (694 bp); DIN4 > RIN4 (723 bp); DIN5 > RIN5 (451 bp). Membrane detection was performed using the Roche kits mentioned above. The images were analyzed using Total Lab100 software (Nonlinear Dynamics, Ltd., Newcastle on Tyne, UK); through software evaluations, bands belonging to different individuals were considered the same truncation variant for differences up to ± 10 bp.

28S rRNA sequence variation

The nucleotide variability of the 28S genes harboring or lacking R2 was analyzed through the amplification, cloning, and sequencing (as above described) of two regions extending from the R2 insertion site to 738 bp upstream and 810 bp downstream (Figure 1). The analysis was performed on eight individuals of the 2004 sample (M2, M3, M4, M6, F24, F25, F27, and F32); three of these specimens lack the R2 full-length element (M4, M6, F32). Sequences have been entered in Genbank, under the accession numbers GU220077–GU220356.

Proportions of nucleotide differences (calculated as mean p-distances within each individual, p-D) and gene diversities (*H*) were calculated for both 28S regions; each value was taken as data point for further elaboration. Two-tailed Student *t*-tests, with equal variance, were computed to assess the significance of differences among the scored variability values assuming that clones harboring R2 are more variable than those lacking the insertion. A second comparison was performed between clones belonging to the individuals with the complete element (M2, M3, F24, F25, and F27) and individuals without the complete element (M4, M6, and F32), both for rDNA units harboring R2 and lacking R2. Finally, a test for selection was performed on both single and pooled datasets using the Tajima's *D* parameter.

Results

The R2 element in *T. cancriformis*

To construct the complete sequence, six clones containing the whole insertion site at the 5′ terminus (5′-TTAA↓GG TAGC-3′; Burke *et al*., 1999; Kojima and Fujiwara, 2005)

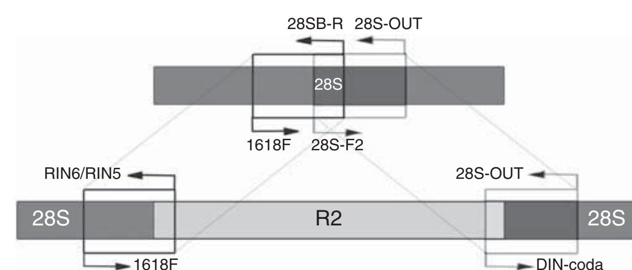

Figure 1 Graphic representation of the regions sequenced for the 28S analyses, with corresponding primer pairs (listed in Table 2). The primer RIN5 was specifically used for individuals M6 and M9, which lack the complete element (see text).

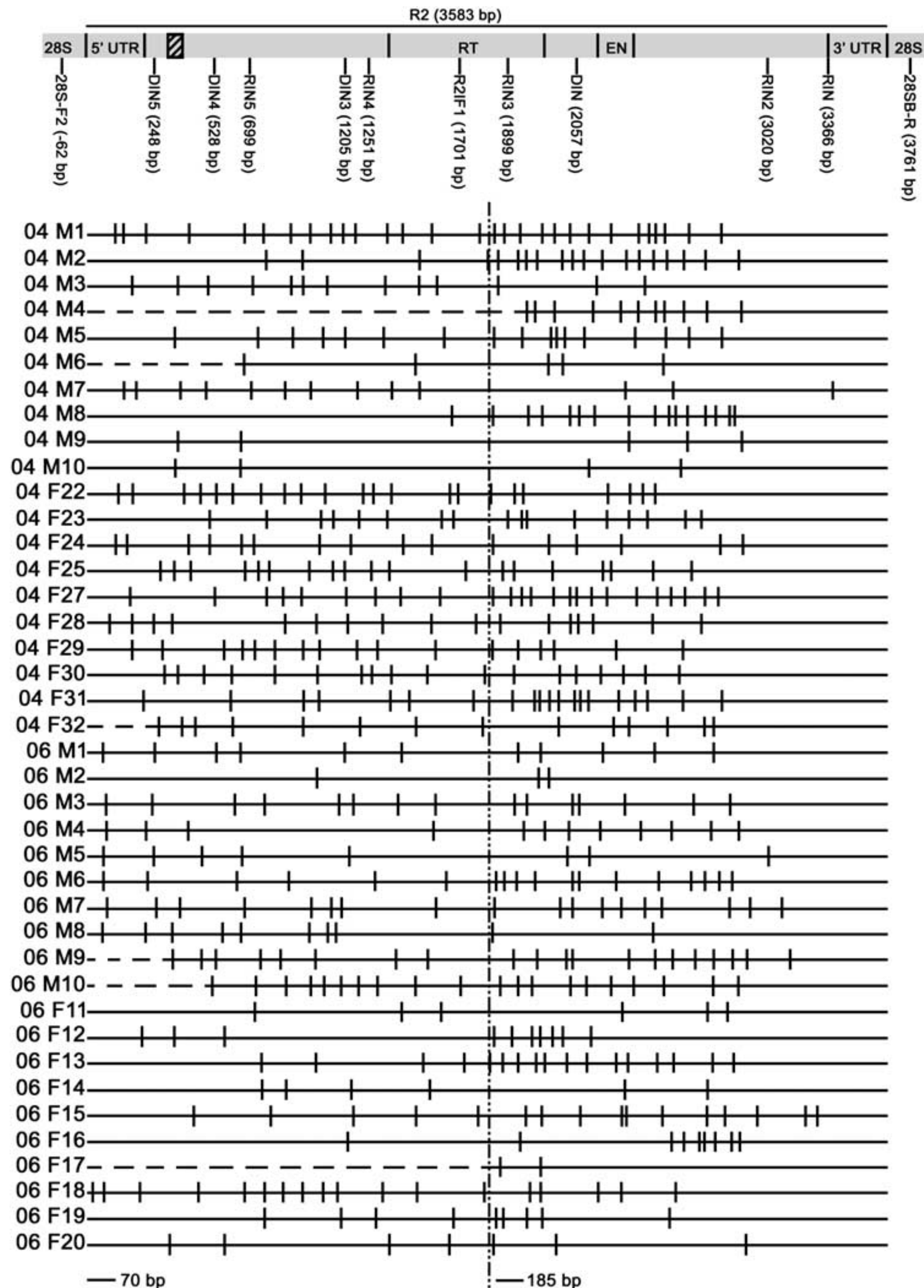

Figure 2 Graphic representation of R2 truncation variants distribution (shown by solid vertical lines) in the 40 *T. cancriformis* individuals. Dashed horizontal lines represent sequences missing in all elements from that individual; the dotted vertical line represents a change in the scale of the *x* axis. A diagram of the R2 element and the primer names/positions with respect to the insertion site are shown at the top. Pattern box indicates the zinc-finger motif; RT, retrotranscriptase domain; EN, endonuclease domain.

were first considered. Other clones showed deletions of the 28S gene that range from 4 to 10 bp.

The sequence of the full-length R2 element here characterized is a consensus of all sequenced fragments obtained by primer walking. It is 3583 bp long (GenBank A.N.: EU854578) and exhibits an A + T content equal to 53%. The sequencing of the R2 5′ end showed a poly-T run of five nucleotides in all analyzed clones; the analysis of the 3′ end revealed a poly-A tail, which is another common feature of R2 mobile elements (Burke *et al.*, 1999; Kojima and Fujiwara, 2005). The R2 sequence contains a 3093 bp long ORF, located between nucleotide 177 and nucleotide 3272, coding for 1031 amino acids. As expected, the ORF is characterized by a reverse transcriptase domain and an endonuclease domain (Figure 2). Moreover, it exhibits a single zinc-finger motif. The comparison between the ORF nucleotide sequences of the R2 elements from *Triops longicaudatus* and *T. cancriformis* shows that 881 out of 1503 bp are variable sites (58%) and they occur at the first codon

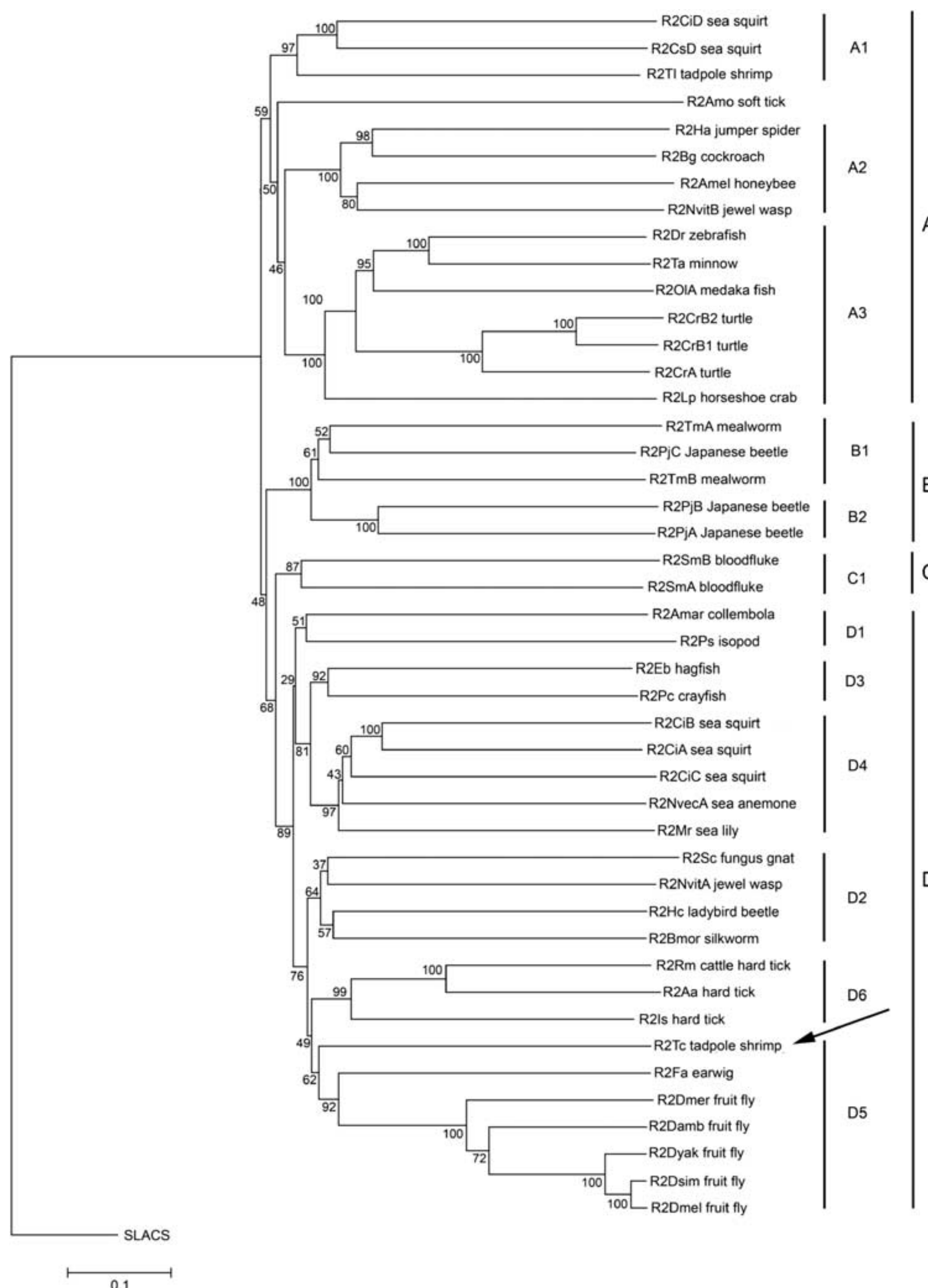

Figure 3 R2 phylogeny inferred by Neighbor Joining method. The number next to each node indicates the bootstrap value, as a percentage of 1000 replicates. Letters indicate clades and subclades (as described in the text); the arrow indicates the position of R2Tc element described here.

position (32.8%) at the second (26.6%) and at the third (40.6%). In total, 346 amino-acid substitutions are scored between the two lineages.

In *T. cancriformis*, R2 occurs at low copy number: only 0.54–5.3% of rDNA units (that constitute the 0.1% of the genome) have R2 insertions.

Phylogenetic analysis

The same terminal branching pattern is observed for all phylogeny estimation methods, even if with different node support values. In the Neighbor Joining dendrogram (Figure 3), clades A, B, C, and D and subclades found by Kojima and Fujiwara (2005) can be recognized with some differences possibly occurring owing to the addition of new sequences. The main variation is given by the presence of a new subclade, named D6, that is composed by R2 elements from three tick species (family Ixodidae): R2Is (*I. scapularis*), R2Aa (*A. americanum*), and R2Bmi (*B. microplus*). This subclade represents the sister branch of the D5 subclade. The element of the fourth tick species (R2Amo, *A. monolakensis*) lies, instead, in clade A, being basal to the A2 and A3 subclades. R2 from *B. germanica* (R2Bg) belongs to the A2 subclade, as sister branch of R2Ha (*Hasarius adansoni*, jumping spider) and

Table 3 Mean sequence variability (p-D), gene diversity (*H*), and Tajima's *D* (per individual and overall) of inserted (R2+) and uninserted (R2−) 28S rDNAs

	p-D		H		*Tajima's* D		
	R2+	*R2−*	*R2+*	*R2−*	*R2+*	*R2−*	*All*
Upstream							
04-M2	0.0014	0.0022	0.533	0.778	−1.741*	−1.873*	−2.316**
04-M3	0.0039	0.0011	0.972	0.533	−1.889*	−1.667	−2.394**
04-M4[a]	0.0027	0.0027	0.972	0.833	−1.822*	1.822*	−2.386**
04-M6[a]	0.0081	0.0054	0.867	0.893	−2.009**	−1.807*	−2.502**
04-F24	0.0036	0.0037	0.972	1.000	−1.795*	−1.756*	−2.338**
04-F25	0.0035	0.0036	1.000	0.917	−1.795*	−1.876*	−2.415**
04-F27	0.0070	0.0116	0.978	1.000	−1.913*	−1.518	−2.275**
04-F32[a]	0.0035	0.0031	0.972	0.917	−1.667	−1.436	−2.275**
Overall	*0.0048*	*0.0042*	*0.945*	*0.869*	*−2.873****	*−2.859****	*−2.932****
Downstream							
04-M2	0.0027	0.0041	0.917	0.978	−1.729	−1.710	−2.252**
04-M3	0.0029	0.0051	0.972	1.000	−1.632	−1.937*	−2.420***
04-M4[a]	0.0015	0.0012	0.800	0.786	−0.447	−1.534	−1.810*
04-M6[a]	0.0067	0.0086	1.000	1.000	−1.211	−1.495*	−2.178**
04-F24	0.0059	0.0037	1.000	0.972	−1.967**	−1.629	−2.444***
04-F25	0.0022	0.0040	0.722	0.833	−1.797*	−1.745*	−2.368**
04-F27	0.0029	0.0042	0.933	1.000	−1.390	−1.642	−2.130**
04-F32[a]	0.0045	0.0027	1.000	1.000	−1.444	−1.421	−2.187**
Overall	*0.0041*	*0.0041*	*0.929*	*0.943*	*−2.832****	*−2.866****	*−2.927****

[a]Samples with no full-length R2 element. *$P<0.05$; **$P<0.01$; ***$P<0.001$.

finally R2NvecA (*N. vectensis*, sea anemone) belongs to subclade D4.

The R2 sequence of *T. cancriformis* here characterized (henceforth called R2Tc) is in a basal position in the D5 subclade, whereas the element belonging to the congeneric *T. longicaudatus* (R2Tl) still lies in the A1 subclade (Kojima and Fujiwara, 2005).

Truncation analysis

The results of Southern blots on R2 truncation patterns are summarized in Table 1 and Figure 2. In the 2004 sample, a total of 318 truncations were detected, ranging from 4 to 28 per individual. In the 2006 sample, 239 variants were scored, ranging from 2 to 20 per individual. Generally speaking, a wide range of truncation profiles has been scored, as each specimen shows its own set of truncations. No ancestral variants have been detected: indeed, there is not any truncation variant present in all individuals; the most widespread one is found in 10 and 7 individuals of 2004 and 2006 collections, respectively. Moreover, six individuals, from both 2004 and 2006 samples (Table 1; Figure 2), presented a set of R2 truncations (ranging from 2 to 20), but did not have the complete element.

28S sequence variability analysis

The nucleotide variability of the 28S gene was studied using two sub-samples: the first one comprised individuals harboring both complete and truncated R2 copies (individuals M2, M3, F24, F25, and F27), whereas the second sub-sample comprised individuals only with truncation variants, but lacking the complete element (individuals M4, M6, and F32; Table 3). For each specimen, sequencing was performed for both R2-inserted and R2-uninserted 28S, upstream and downstream of the insertion site (Figure 1). From 6 to 10 sequences per individual were obtained for 28S rDNAs with or without R2 insertions, for both upstream and downstream regions.

A total of 147 28S sequences 738 bp long were obtained for the upstream region: 76 carrying the element (hence R2+) and 71 without it (hence R2−). In the R2+ dataset, 110 polymorphic sites were found resulting in 56 alleles; the mean sequence diversity within individual varies widely, from 0.0014 to 0.0081 (overall = 0.0048), as well as the gene diversity, from 0.533 to 1.000 (overall = 0.945; Table 3). Sequences of the R2− dataset show 93 polymorphic sites and are, on average, slightly less variable: the overall sequence variability is equal to 0.0042 (varying from 0.0011 to 0.0116) and the overall gene diversity is 0.869 (ranging from 0.533 to 1.00; Table 3). It is to be noted that the 04-F27 individual shows an R2− mean p-distance value of 0.0116, significantly higher than the population average (0.0042). Grubb's test for outliers resulted significant for that value ($P<0.05$); therefore, it has been excluded from subsequent analyses.

For the downstream region, 133 28S sequences 810 bp long were obtained: 63 for the R2+ dataset and 70 for the R2− one. In the former alignment, 83 sites were found variable, whereas in the latter, 95 were polymorphic. Also in this region, sequence and gene diversities vary widely: 0.0015–0.0067 and 0.800–1.000, respectively, for the R2+ dataset; 0.0012–0.0086 and 0.786–1.000, respectively, for the R2− one (Table 3). Overall, R2+ and R2− values for both parameters are almost the same: sequence variability is equal to 0.0041 for both dataset; gene diversity results 0.929 and 0.943, respectively. Interindividual sequence variabilities are reported in Supplementary Tables S1 and S2.

Student *t*-tests performed on both sequence and gene diversity measures did not show any significant comparison between R2+ and R2− datasets. R2 presence does

not seem to interfere with the 28S homogenization process. Moreover, tests conducted between individuals carrying the complete element and those without it again did not show any significant comparison, for both R2+ and R2− datatsets and for both upstream and downstream regions.

Tajima's *D* test performed on single R2+ and R2− alignments, as well as for pooled datasets, rejects the neutrality hypothesis in the majority of trials, especially for the upstream region; moreover, all values, whether significant or not, are negative (Table 3).

Discussion

The R2 element from *T. cancriformis* well matches general features of this kind of transposable element. So, a single ORF with a central reverse transcriptase domain is present, as in type II non-LTR retrotransposons (Eickbush and Jamburuthugoda, 2008). In addition, its insertion site within the *T. cancriformis* 28S gene is the same as in all studied organisms (Burke *et al*., 1999; Kojima and Fujiwara, 2005) save ticks, which show two nucleotide substitutions within the normal target sequence (Bunikis and Barbour, 2005). In *T. cancriformis*, a four base deletion (TTAA) was found in R2-inserted 28S, whereas in the *Drosophila* genus 28S deletions are larger. On the other hand, the 3′ end of R2Tc is congruent with that of *Drosophila* spp., with the typical poly-A tail and the deletion of the two Gs at the insertion site (George *et al*., 1995; Burke *et al*., 1999; Pérez-Gonzalez and Eickbush, 2001). These features are determined by the *target primed reversed transcription* mechanism, in the phase of DNA cleavage and cDNA synthesis (Christensen *et al*., 2006).

R2 phylogeny

Earlier studies outlined an important aspect of R2 evolution: with only few exceptions, the R2 and host phylogenies do not overlap. Two hypotheses have been put forward to explain this pattern: in the first, vertical inheritance of the element can be followed by lineage extinction or diversification in certain groups; the second hypothesis assumes the horizontal transfer of R2 between species. In a recent survey, the former has been shown as the most likely explanation (Kojima and Fujiwara, 2005), so that the incongruence between host and R2 phylogeny can be explained, almost totally, by high rates of diversification of the element and not by horizontal transfer between species. The matter is very intriguing, because deep nodes of R2 phylogeny are consistent with structural features (number of zinc-finger motifs at the N-terminus; Kojima and Fujiwara, 2005) and, if we assume a vertical inheritance followed by diversification, some still-unknown factor should underlie this consistency. Obviously, the two hypotheses (vertical vs horizontal transmission) may not be mutually exclusive.

As expected from the occurrence of one zinc-finger motif, R2Tc falls in the D clade (Kojima and Fujiwara, 2005) and does not cluster with the *T. longicaudatus* element, which lies within the group of elements with three zinc-finger motifs. Actually, this is not currently verifiable because the only available *T. longicaudatus* R2 element is not complete. Beside structural features, the two notostracan R2 lineages are also very different both at the nucleotide (58%) and amino-acid (69%) levels. Owing to its observed high rate of retrotransposition, R2 sequences are subject to a high diversification leading to the occurrence of multiple lineages within the same species and/or elimination of some lineages owing to competition for the limited number of insertion sites (Pérez-Gonzalez and Eickbush, 2001). Moreover, R2s are ancient components of the animal genome, their presence dating back at least to the splitting of cnidarians and bilaterians (Kojima *et al*., 2006). The antiquity and the evolutionary dynamics of R2 may explain, therefore, the lack of correlation between its phylogeny and that of the host species. In this view, the divergence of the two tadpole shrimp R2 sequences might be the result of such an evolutionary process.

Truncation analysis

R2 truncation analyses have been first conducted on laboratory stocks of *Drosophila melanogaster* and *D. simulans*. In the former, the variants distribution has been found, to some extent, well conserved, with ancestral-truncated variants being shared by individuals both within and between isofemale lines. However, some lines of *D. simulans* show decidedly higher R2 activity, producing less conserved truncation profiles (Pérez-Gonzalez and Eickbush, 2001, 2002; Perez-Gonzalez *et al*., 2003; Zhang and Eickbush, 2005). A recent survey on natural populations of *D. simulans* showed a high turnover rate, each individual carrying a specific collection of R2 truncations (Zhou and Eickbush, 2009). The high incidence of R2 insertions in *D. simulans* is correlated with a high rate of variant elimination and a lower number of inserted 28S, explanable as due to its retrotransposition creating large deletions in adjacent rDNA units, thus eliminating a number of R2 variants (Zhang *et al*., 2008). The dynamics of R2 in *T. cancriformis* is in line with these observations as (i) individuals from the same and/or different samples show very different truncation profiles, (ii) there are not ancestral variants shared by all individuals, and (iii) the percentage of rDNA units with insertions is very low (∼0.5–5%). However, a peculiarity related to the R2 elimination occurs: six tadpole shrimps show truncated variants, but not the complete element. Therefore, an active R2 is lacking in their genomes.

In *D. simulans*, inserted 28S are hypothesized to be eliminated by the transposition of active R2 elements, whereas genomic turnover mechanisms tend to replace deleted rDNA units with new ones for the maintenance of the ribosomal locus functionality. This, however, also creates new niches for the R2 element, which can remain active (Eickbush and Eickbush, 2007; Zhang *et al*., 2008). On the other hand, as a consequence of transposition-mediated deletions, the loss of R2 variants, either complete or truncated, might be dramatic: the element copy number can be reduced to very few copies (for example, a single 28S rDNA with an insertion). Moreover, genomic turnover mechanisms acting on rDNA locus might eliminate all 28S carrying insertions. Therefore, a process such as transposition-mediated deletion, together with unequal DNA exchanges acting on the few 28S units carrying the complete R2 elements, can explain why these variants are lacking in the 15% of the *T. cancriformis* assayed. Once

the full-length copy is deleted, new insertions cannot occur and the remaining truncations would be progressively eliminated by subsequent rounds of genomic turnover mechanisms (Dover, 2002).

The loss of the R2 element from a genome is the first step toward the extinction in a given population/species, possibly leading to unclear phylogenetic patterns (see above); it is, therefore, essential to understand how this mechanism proceeds. The elimination of R2 through the interplay between transposition-mediated deletions and genomic turnover mechanisms might give a clue to the process, but how the absence of R2 can be maintained in a population is still a further, open question. In a gonochoric population, as the studied tadpole shrimp samples are, outcrosses between individuals without the complete element and individuals carrying a functional R2 will very likely result in an offspring with active (complete) elements. This would mean that the extinction of an R2 lineage in a given population, in absence of other factors, is a very unlikely event, albeit it has been already shown (Jakubczak *et al*., 1991). The occurrence of some selective advantages at one particular stage of the population/species life history, leading to a preferential survival of the individuals lacking R2, can be suggested. However, we neither have any direct evidence, nor it is possible to draw similar conclusions from the datasets presented so far. Alternatively, as discussed in the earlier paragraph, the peculiar dynamics of this non-LTR retrotransposon may generate the presence of multiple R2 lineages in the same genome, competing for the limited insertion sites (Pérez-Gonzalez and Eickbush, 2001). Thus, the possibility of an R2 lineage's replacement, which would prevent the annealing of designed primers, cannot be excluded.

28S variation analysis

The presence of R2 (either truncated or not) within a 28S gene can influence the ribosomal sequence homogeneity in two ways: (i) a large insertion (kilobases long) may interfere with recombination, preventing the pairing of inserted with uninserted rDNA repeats; (ii) once R2 is inserted within a 28S sequence, the ribosomal gene becomes a pseudogene and may freely accumulate mutations. In both instances, R2 insertion can hinder concerted evolution, basically avoiding the homogenization process. It can, therefore, be expected that inserted sequences would be more variable than the uninserted ones. Here, the presence of R2 does not impact on 28S sequence and gene diversity. Indeed, comparisons between R2+ and R2− 28S genes are not significant, showing very close variability values. This is in line with data on *Drosophila*, in which inserted and uninserted 28S are identical: this has been explained through the rapid elimination of new insertions (Eickbush and Eickbush, 2007). Differently, 28S rRNA genes carrying the *Pokey* element in *D. pulex* accumulate more mutations than those without the insertion: the authors explain this contrasting pattern as the results of the long persistence (or even the spreading) of some *Pokey* insertions preventing ribosomal unit recombination and homogenization (Glass *et al*., 2008). The R2 turnover suggested for *T. cancriformis* would not allow this dynamics, as the newly transposed variants would be rapidly eliminated. As argued by Glass *et al*. (2008), recently generated 28S-inserted copies would be indistinguishable from those that never experienced the element insertion and, because of their quick elimination, they do not significantly alter the rDNA homogenization level.

The generally low variability observed in this analysis suggests a quite efficient process of sequence conservation and the hypothesis of neutrality has been rejected in several instances. Interestingly, the majority of significant Tajima's Ds can be observed in the region upstream of the insertion site, whereas in only few instances this has been shown in the downstream region. The downstream region here characterized is homologous to that described by Glass *et al*. (2008) in *D. pulex* in which the same, non-significant values have been observed: this may indicate that this region undergoes neutral evolution.

Generally speaking, all Tajima's D values observed here are negative, evidencing an excess of low frequency polymorphisms: this can be either the results of purifying selection (that can be expected, of course) or caused by a recent expansion of new 28S variants. Measures of gene diversity are consistent with the latter scenario as the higher values obtained are expected when there are several alleles none of which reaching very high frequency. This well reconciles with the R2 turnover observed here: as recalled in the earlier paragraph, as for any retrotransposition event large rDNA units deletions occur (Zhang *et al*., 2008), a compensatory replacement of new 28S variant is necessary for proper functionality. Multiple cycles of rDNA unit gains and losses would boost their turnover, leading to a quite homogeneous array (as rDNA units are very recently duplicated), but at the same time let spread several single point mutations throughout the array.

It would be interesting to investigate if, in the absence of R2, the same pattern of sequence and gene diversity can be achieved. The three individuals without the complete (=active) R2 element show no difference in comparison with those having the full-length retrotransposon; however, as they were sampled from a random mating population, it is unlikely that subsequent generations will lack the active element.

The R2/rDNA 'interplay' can be interpreted as a reciprocal advantage: new niches for R2, more efficient homogenization of rDNA units. Does this dynamics bring further advantage to the host? The turnover of rDNA units within the array, in which the retrotransposition occurs, may lead to a greater variance in the proportion of functional/defective 28S rRNA genes between individuals. In an evolutionary perspective, this would result in more opportunities for natural selection to operate on the host.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr Dani Boix, Dr Jordi Sala (Girona University), and Dr Loris Mularoni (IMIM, Barcelona) for collecting the specimens, and Professor Haruhiko Fujiwara for supplying part of the R2 protein alignment. This work was funded by Donazione Canziani and Fondazione del Monte di Bologna e Ravenna grants. We also thank two anonymous referees for their constructive

criticism that substantially improved this paper. This paper is dedicated to Professor Franca Scanabissi: from her reproductive and ultrastructural point of view, she has beautifully introduced us to the *Triops* (and other Branchiopoda) world, her collaboration having always strongly supported the research group.

References

Aksoy S, Williams S, Chang S, Richards FF (1990). SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucleic Acids Res* **18**: 785–792.
Bunikis J, Barbour AG (2005). Ticks have R2 retrotransposons but not the consensus transposon target site of other arthropods. *Insect Mol Biol* **14**: 465–474.
Burke WD, Eickbush DG, Xiong Y, Jakubczak J, Eickbush TH (1993). Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. *Mol Biol Evol* **10**: 163–185.
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.
Christensen SM, Ye J, Eickbush TH (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.
Dover GA (1982). Molecular drive, a cohesive model of species evolution. *Nature* **299**: 111–117.
Dover GA (2002). Molecular drive. *Trends Genet* **18**: 587–589.
Eickbush TH, Burke WD, Eickbush DG, Lathe III WC (1997). Evolution of R1 and R2 in the rDNA units of the genus *Drosophila*. *Genetica* **100**: 49–61.
Eickbush TH, Eickbush DG (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**: 477–485.
Eickbush TH, Jamburuthugoda VK (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* **134**: 221–234.
Fisher DC (1990). Rate of evolution in living fossils. In: Briggs DEG, Krauther PR (eds). *Paleobiology*. Blackwell Scientific: London. pp 152–159.
George JA, Burke WD, Eickbush TH (1995). Analysis of the 5′ junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. *Genetics* **142**: 853–863.
Glass SK, Moszczynska A, Crease TJ (2008). The effect of transposon Pokey insertions on sequence variation in the 28S rRNA gene of *Daphnia pulex*. *Genome* **51**: 988–1000.
Halanych KM (2004). The new view of animal phylogeny. *Annu Rev Ecol Evol Syst* **35**: 229–256.
Hillis DM, Dixon MT (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* **66**: 411–453.
Jakubczak JL, Burke WD, Eickbush TH (1991). Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc Natl Acad Sci USA* **88**: 3295–3299.
Kagramanova AS, Kapelinskaya TV, Korolev AL, Mukha DV (2007). R1 and R2 retrotransposons of German cockroach *Blattella germanica*: a comparative study of 5′-truncated copies integrated into the genome. *Mol Biol* **41**: 546–553.
Kojima KK, Fujiwara H (2004). Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol Biol Evol* **21**: 207–217.
Kojima KK, Fujiwara H (2005). Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22**: 2157–2165.
Kojima KK, Kuma K, Toh H, Fujiwara H (2006). Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. *Mol Biol Evol* **23**: 1984–1993.
Korn M, Marrone F, Pérez-Bote JL, Machado M, Cristo M, Cancela da Fonseca L *et al*. (2006). Sister species within the *Triops cancriformis* lineage (Crustacea, Notostraca). *Zool Scr* **35**: 301–322.
Malik HS, Burke WD, Eickbush TH (1999). The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16**: 793–805.
Mallatt JM, Garey JR, Shultz JW (2004). Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Mol Phylogenet Evol* **31**: 178–191.
Mantovani B, Cesari M, Luchetti A, Scanabissi F (2008). Mitochondrial and nuclear DNA variability in the living fossil *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca). *Heredity* **100**: 496–505.
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.
Nei M, Rooney AP (2005). Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**: 121–152.
Penton EH, Crease TJ (2004). Evolution of the transposable element *Pokey* in the ribosomal DNA of species in the subgenus *Daphnia* (Crustacea: Cladocera). *Mol Biol Evol* **21**: 1727–1739.
Perez-Gonzalez CE, Burke WD, Eickbush TH (2003). R1 and R2 retrotransposition and elimination from the rDNA loci of the X and Y chromosomes of *Drosophila melanogaster*. *Genetics* **165**: 675–685.
Perez-Gonzalez CE, Eickbush TH (2001). Dynamics of R1 and R2 elements in the rDNA locus of *Drosophila simulans*. *Genetics* **158**: 1557–1567.
Pérez-Gonzalez CE, Eickbush TH (2002). Rates of R1 and R2 retrotransposition and elimination from the rDNA locus of *Drosophila melanogaster*. *Genetics* **162**: 799–811.
Rasband WS (1997–2007) ImageJ. U.S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/.
Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
Swofford DL (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates: Sunderland, Massachusetts.
Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
Zhang X, Eickbush TH (2005). Characterization of active R2 retrotransposition in the rDNA locus of *Drosophila simulans*. *Genetics* **170**: 195–205.
Zhang X, Zhou J, Eickbush TH (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, Eickbush TH (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.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)