A deep-branching clade of retrovirus-like retrotransposons in bdelloid rotifers

Running title: Rotifer LTR retrotransposons

Keywords: LTR retrotransposons; asexual reproduction; reverse transcriptase; integrase; gag; pol; env

Eugene A. Gladyshev[1], Matthew Meselson[1,2], and Irina R. Arkhipova[1,2*]

[1]*Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA;*

[2]*Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA*

Abbreviations: aa – amino acid(s); bp - base pair(s); IN – integrase; kb – kilobase(s); LTR – long terminal repeat; Myr – million years; ORF – open reading frame; RT – reverse transcriptase; TE – transposable element; TM – transmembrane; UTR – untranslated region.

Address for correspondence: *Dr. Irina Arkhipova, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

Tel. (617) 495-7899

Fax: (617) 496-2444

E-mail: arkhipov@fas.harvard.edu

Abstract

Rotifers of class Bdelloidea, a group of aquatic invertebrates in which males and meiosis have never been documented, are also unusual in their lack of multicopy LINE-like and gypsy-like retrotransposons, groups inhabiting the genomes of nearly all other metazoans. Bdelloids do contain numerous DNA transposons, both intact and decayed, and domesticated *Penelope*-like retroelements *Athena*, concentrated at telomeric regions. Here we describe two LTR retrotransposons, each found at low copy number in a different bdelloid species, which define a clade different from previously known clades of LTR retrotransposons. Like bdelloid DNA transposons and *Athena*, these elements have been found preferentially in telomeric regions. Unlike bdelloid DNA transposons, many of which are decayed, the newly described elements, named *Vesta* and *Juno*, inhabiting the genomes of *Philodina roseola* and *Adineta vaga*, respectively, appear to be intact and to represent recent insertions, possibly from an exogenous source. We describe the retrovirus-like structure of the new elements, containing *gag*, *pol*, and *env*-like open reading frames, and discuss their possible origins, transmission, and behavior in bdelloid genomes.

1. Introduction

Rotifers of the class Bdelloidea, comprising some 370 described species, are microscopic freshwater invertebrates of worldwide distribution. The class is of particular interest because its apparent asexuality challenges the widely held view that sexual reproduction is essential for long-term evolutionary success and that its abandonment leads to early extinction (Bell, 1982; Mark Welch and Meselson, 2000; Normark *et al*., 2003). Although there are many hypotheses, there is no general agreement as to what accounts for the nearly universal occurrence of sexual reproduction in animals and plants and the relatively early extinction of asexual lineages (Kondrashov, 1993; Barton and Charlesworth, 1998; Otto and Lenormand, 2002). According to one of the hypotheses, a central factor in the maintenance of sexual reproduction is its role in limiting the increase of parasitic mobile genetic elements, which, if unchecked, would drive their hosts to extinction (M.M., in Arkhipova *et al*., 1995; Arkhipova and Meselson, 2000, 2005a).

In order to investigate this hypothesis and the possible connections between long-term asexuality and transposable elements (TE), we initiated studies of TE content in several bdelloid species. We first performed a PCR screen with highly degenerate primers for the most widespread TE superfamilies, namely LINE-like and gypsy-like retrotransposons and mariner/Tc-like DNA transposons (Arkhipova and Meselson, 2000). No LINE-like or gypsy-like reverse transcriptases were detected in any of the 5 bdelloid species tested, although LINEs were easily detected in 39 and gypsies in 35 other species representing 23 animal phyla, while mariner/Tc-like transposases were found in bdelloids and exhibited their characteristically patchy distribution among other species. In a subsequent survey of DNA transposons, the pattern of mariner evolution was found to be consistent with repeated horizontal transfers from an exogenous

source, followed by limited intragenomic proliferation and eventual decay (Arkhipova and Meselson, 2005b). These and other types of DNA transposons (hAT, piggyBac, helitron, foldback), often inserted into each other, were found to be largely concentrated near chromosome ends, associated with *Athena* retroelements, but notably lacking in the more proximal, gene-rich regions thus far sequenced. *Athena* retroelements cannot be assigned to either LTR or non-LTR retrotransposons, but instead belong to *Penelope*-like elements, a recently described class of eukaryotic retroelements forming a sister clade to telomerase reverse transcriptases (Arkhipova *et al*., 2003). The movement of these intron-containing retroelements is apparently confined to chromosome ends, where they may act in a backup pathway for telomere elongation (E.G. and I.A., submitted).

During the sequencing of a telomere-associated *Athena*-containing cosmid from the bdelloid *Philodina roseola* (Arkhipova and Meselson, 2005b), we encountered an LTR retroelement that appeared to be more closely related to chromoviruses (Kordis, 2005) than to any other known group of LTR retrotransposons. We now find that, together with a related but distinct element from *Adineta vaga*, a bdelloid species from a family that separated from the Philodinidae tens of millions of years ago, these new elements, designated *Vesta* and *Juno*, respectively, define a deep-branching clade of LTR retrotransposons, appearing as a sister clade to chromoviruses. Within each species, the copies of these elements are only slightly divergent and several copies have been localized to regions near telomeres. Present at very low copy number and apparently intact, *Vesta* and *Juno* may have entered the respective bdelloid genomes only recently, their horizontal entry being facilitated by a putative *env* region from an unknown source.

2. Materials and Methods

2.1. DNA manipulations.

The genomic cosmid library of *P. roseola* embryo DNA (*ca*. 5x coverage) prepared by partial *Sau*3AI digestion (Mark Welch *et al*., 2004) was screened with a ^{32}P-labeled *Vesta* probe (Fig. 1a) and a PCR-generated *Athena* probe (Arkhipova *et al*., 2003). The *A. vaga* genomic fosmid library (*ca*. 6x coverage) prepared from randomly sheared embryo DNA (J. Hur, Ph.D. dissertation, Harvard University) was screened with a ^{32}P-labeled $(TGTGGG)_4$ probe and a *Juno* probe (Fig. 1b). Library screening and isolation of cosmid and fosmid clones were performed according to instructions from manufacturers of library production kits (Stratagene and Epicentre Technologies). Hybridizing cosmids and fosmids were analyzed by restriction enzyme digestion and by direct sequencing with standard T3, T7 and pCC1Fos primers and with primers used to generate the probes. Fosmid/cosmid inserts selected for complete sequencing were sheared by sonication, subcloned into pBluescript II SK- (Stratagene), and sequenced with T3 and T7 primers using BigDye Terminator v3.1 (Applied Biosystems) and ABI 3730 XL capillary sequencers at the high-throughput W.M. Keck Ecological and Evolutionary Genetics Facility at the MBL. Four additional *Juno* copies were sequenced by primer walking with a set of 12 custom oligonucleotides.

2.2. Bioinformatics.

Sequences were assembled with Phred/Phrap/Consed (www.phrap.org), and aligned with the AlignX program, based on the ClustalW algorithm, from the VectorNTI suite 7 (InforMax). Alignments are available from the corresponding author upon request. Ks and Ka/Ks ratios were calculated by the DIVERGE program of the GCG Wisconsin package (Accelrys), which uses the method of Pamilo and Bianchi (1993). Phylogenetic analysis was performed with MEGA 3.1 (Kumar *et al*., 2004) by neighbor-joining (pairwise

deletion; Poisson correction model; uniform among-site rate variation), and with MRBAYES 3.1.1 (Ronquist and Huelsenbeck, 2003) using the mixed amino acid substitution model, invariable plus gamma among-site rate variation, 10^6 generations, each 100th tree sampled, and the first 200 trees discarded as burn-in. Sequences obtained in this study were deposited in GenBank under accession numbers **DQ985390**-**DQ985395**.

3. Results

3.1. Structural organization.

The two LTR retrotransposons characterized in this study share a highly similar overall organization, with three open reading frames (ORFs) in the same order in both elements (Fig. 1). The *P. roseola* element, *Vesta*, was found on an *Athena*-containing telomeric cosmid, while the *A. vaga* element, *Juno*, was initially found on a fosmid selected from a genomic library by hybridization with *A. vaga* telomeric repeats. Additional *Juno* copies were isolated by probing the *A. vaga* genomic library with an internal fragment (Fig. 1). The cloned copy of *Vesta* is 7981 bp in length, while *Juno* is slightly longer (consensus 8136 bp), mainly because of a longer ORF3. Both elements generate a 5-bp non-specific target-site duplication (TSD) upon insertion. A comparative description of the functionally significant regions of the two elements is given below.

3.1.1. LTRs and noncoding regions.

The LTRs of *Vesta* and *Juno* are very compact, 343 and 283 bp in length, respectively (Fig. 1; Table 1). Both begin with the canonical TG dinucleotide and have short terminal inverted repeats at their boundaries, TGTTA...TAACA in *Vesta* and TgTTA...TAAtA in *Juno*. While such short LTRs are not likely to accommodate enhancer sequences, the elements of a basal eukaryotic promoter can easily be discerned: transcription probably

starts within a canonical initiator (*Inr*) element, TCAGT (Smale and Baltimore, 1989; Arkhipova, 1995), with *Juno* also containing a TATATA motif 25-30 bp upstream of *Inr*. Both LTRs also carry a polyadenylation signal, AATAAA, as required for production of a putative transcript with a short terminal redundancy (R) and formation of the characteristic U3-R-U5 structure of retrotransposon LTRs (Arkhipova *et al*., 1986). The TGG of the putative primer-binding site, presumed to be complementary to the 3' CCA of an unidentified host tRNA, is separated from the left LTR by a dinucleotide (as in retroviruses) in *Juno*, and by 3 bp in *Vesta* (Table 1). Although polypurine tracts (PPT) adjacent to the right LTRs are not very pronounced (Table 1), functional studies of *Ty1* PPT in yeast demonstrate that a PPT with 6 purines out of 9 nucleotides is sufficient for plus-strand priming (Heyman *et al*., 1995). The 5' and 3' UTRs are also very compact (Tables 1, 2). The compact nature and low sequence complexity of the LTR and UTR regions indicates that *Vesta* and *Juno* likely possess a basal promoter but have not acquired any extended regulatory regions from the host, as often happens during host-TE co-evolution (*e.g*. Wilson *et al*., 1998; Spana *et al*., 1988). Finally, the non-coding region between *Juno* ORF2 and ORF3 exhibits hyper-variability in comparison with the rest of the element (Table 2).

3.1.2. Coding regions.

3.1.2.1. Gag

In most retrovirus-like elements ORF1, corresponding to the *gag* region, codes for proteins that bind to viral template RNA to form the nucleoprotein core particle (Coffin *et al*., 1997). Both *Vesta* and *Juno* ORF1s contain regions homologous to the major capsid (CA) (Retrotrans_gag, pfam03732) and nucleocapsid (NC) proteins (zf-CCHC, pfam00098) (Fig. 2). The *Vesta* NC protein contains one Zn knuckle motif with the characteristic retrovirus-like spacing $CX_2CX_4HX_4C$, expected to play a role in nucleic

acid binding, and the *Juno* NC protein contains two such motifs (Fig. 2, top). The *gag* region of retroviruses/retrotransposons is normally connected to the *pol* region *via* a translational frameshift (Coffin *et al*., 1997). Both *Juno* and *Vesta* exhibit this characteristic –1 frameshift, and apparently use the same mechanism to regulate *pol* expression.

3.1.2.2. Pol

ORF2, the *pol* region, includes domains coding for protease, reverse transcriptase, RNase H, and integrase (PR-RT-RH-IN), in that order, as is characteristic of the Ty3/gypsy-like LTR retrotransposons, the Metaviridae (Boeke *et al*., 2004). A typical retroviral aspartyl protease domain (RVP, pfam00077), found in both elements, has the potential for processing of a gag-pol polyprotein. The reverse transcriptase (RT) domain (RVT, pfam00078), although bearing sufficient similarity to other Metaviridae RTs, differs from them in one of its most conserved motifs, RT4(B): in the nearly universal PFG (or PYG), the aromatic residue in the middle is replaced in *Vesta* by a methionine (which is found in this position only in *Penelope*-like elements and hepadnaviruses), and in *Juno* by a glutamine (typical of retroviruses). The deviant sequences of the most highly conserved RT motifs apparently precluded earlier detection of these retrotransposons in our nested PCR-based assays. The RNase H domain, which is needed for removal of the template RNA within the cytoplasmic RNP particle and for generation of the plus-strand primer, is present in all LTR retrotransposons, and its presence in *Vesta* and *Juno* indicates, together with the presence of the *gag* domain, that these elements undergo reverse transcription in a cytoplasmic RNP particle. Finally, the IN domain (RVE, pfam00665) displays all the characteristics of a retroviral/retrotransposon integrase, including a HH-CC Zn finger and the DDE catalytic domain. The C-terminal GPF/Y motif (Malik and Eickbush, 1999) is clearly evident in *Juno*, but exhibits signs of degeneration

in the cloned copy of *Vesta* (Fig. 2, bottom). This domain is an optional component of the integrase enzyme and is not present in early-branching LTR retrotransposons (Ty1/copia, BEL, DIRS), but was acquired by retroviruses and late-branching Metaviridae, with subsequent loss in many retroviruses and in the gypsy superfamily, also known as insect errantiviruses (Boeke *et al*., 1999).

3.1.2.3. Env

One of the most interesting questions posed by the three-ORF structure of *Vesta* and *Juno* is whether they represent invertebrate retroviruses, *i.e*. whether their ORF3s correspond to the retroviral *env* region and confer the capacity for entering the cell in a manner resembling retroviral infection. The *env* region of retroviruses normally codes for two proteins, a surface protein (SU) binding a host receptor and a transmembrane protein (TM), which are cleaved from the precursor by a host endopeptidase (Coffin *et al*., 1997). Six lineages of LTR retrotransposons are known to have independently acquired such an ORF3, and in three cases a viral source could be identified (Malik *et al*., 2000). Although we could not identify a source for the *env* regions of *Vesta* or *Juno* using the same methods, this is not surprising, since sequence information regarding viruses of invertebrates from aquatic habitats is extremely limited (see below).

Of the expected amino acid sequence motifs characteristic of retroviral *env* proteins (Coffin *et al*., 1997; Misseri *et al*. 2003), we could identify transmembrane (TM) regions (three in *Vesta*, one or two in *Juno*), several N-linked glycosylation sites, a few cysteines that could participate in formation of disulfide bridges, and a putative furin-like protease cleavage site (Fig. 3). Interestingly, TM regions are detected only if ORF3 is extended beyond the stop codon/frameshift, either *via* splicing of a predicted intron, or *via* bypassing a frameshift occurring in a T-rich region, *e.g*. by ribosomal frameshifting.

Similar TM-containing ORF3 extensions in a different reading frame were observed in several endogenous plant retroviruses (Wright and Voytas, 2002). It is conceivable that expression of the functional *env* protein was taking place in the pre-bdelloid hosts, but ceased after entering bdelloid genomes.

3.2. Copy numbers.

Several approaches were used to estimate copy numbers. For *Vesta*, we performed Southern blot hybridization using a probe spanning the *gag-pol* junction (Fig. 1). Up to four hybridizing bands were seen in restriction digests of *P. roseola* genomic DNA (Fig. 4 A), while hybridization at low stringency to genomic DNA from *A. vaga* and from another bdelloid species, *Habrotrocha constricta*, gave no signal (not shown). An independent estimate of 4-5 copies was obtained by probing the *P. roseola* genomic cosmid library with the same *Vesta* fragment and comparison of the number of hybridizing cosmids with the number of cosmids hybridizing under the same conditions to the *hsp82* gene, known to be present in *P. roseola* in four copies, with pairwise divergence between two members of a lineage 4-5% and between members of different lineages up to 47% (Mark Welch *et al*., 2004). Finally, fluorescent *in situ* hybridization with the *Vesta*-containing cosmid (which also contains *Athena* elements, mariner/Tc-like DNA transposons, and a helitron, but no recognizable structural genes in the 40 kb insert) to *P. roseola* embryo nuclei gave 4-6 hybridization signals per nucleus, localized near telomeres (E.G. and I.A., submitted).

For *Juno*, copy number was estimated by probing the *A. vaga* genomic fosmid library with a *Juno* probe spanning the highly conserved RT motifs 4 through 7 of the *pol* region (Fig. 1). All hybridizing fosmids were isolated, and the sequences of *Juno* copies and 0.6-0.7 kb of adjacent flanking regions were determined by primer walking. The 26

hybridizing fosmids obtained from this exhaustive screen were found to correspond to only 5 distinct copies, each with different flanking sequences. The screen was done under conditions in which approximately the same number of fosmids hybridize to a probe for the *hsp82* gene, of which, as in *P. roseola*, there are four copies (J. Hur, personal communication). Copy number estimates from Southern blotting experiments with the same *Juno* probe agree with the estimates obtained from library screening, yielding four strongly hybridizing and one weakly hybridizing band (Fig. 4 B).

3.3. Divergence.

Each of the three ORFs in each of the five *Juno* copies appears intact, even in the most divergent copies (*Juno1* and *Juno4*). Total nucleotide divergence between *Juno* copies in all comparisons is less than 2%, if the hypervariable region between ORF2 and ORF3 is excluded. For each ORF, values of synonymous divergence (Ks) and the ratio of non-synonymous to synonymous divergence (Ka/Ks) are given in Table 3. The excess of synonymous divergence for ORF1 and ORF2 is evidence of purifying selection, while for ORF3 Ka/Ks is close to unity. Comparison between LTRs of different *Juno* copies reveals the identity of LTRs in *Juno2* and *Juno3*, which differ from the LTR of *Juno5* by one substitution and two 1-bp slippages in oligo(T) stretches; *Juno4* LTR differs from them by two more T slippages, and the most divergent LTRs, those of *Juno1,* differ from LTRs of *Juno2-5* by a substitution, 3 slippages involving 2, 4, and 7 bp, and a 7-bp insertion. Differences apparently caused by replication slippage also prevail in the hypervariable region between ORF2 and ORF3, and within the C-terminal part of ORF3, including the presumptive intron. Five *Vesta*-containing cosmids selected from the *P. roseola* genomic library, when sequenced with several internal primers, did not reveal any differences from the fully sequenced copy, although we do not know how many different copies these cosmids represent.

Divergence between LTRs of a given element has been taken as a measure of the time since the element last underwent transposition, since both LTRs are copied from the same regions of the RNA template during this process (R and U5 regions of each LTR are copied from the 5' RNA end, and U3 – from the 3' end). On this view, the identity of left and right LTRs (except for *Juno4*) implies that *Vesta* and *Juno* transposed recently. Taking the Drosophila synonymous substitution rate of 0.016 substitutions per site per Myr (Li, 1997), such identity between LTRs of *ca* 300 bp suggests that the time since the most recent transposition of any element may be no more than a few hundred thousand years. In contrast, if it is assumed that all five copies of *Juno* derive from a single invading copy, its arrival would have occurred several million years ago, as estimated from Ks values between the most divergent copies (*Juno1,4*) and the three more similar copies. Regardless of the particular value taken for the rate of synonymous substitution, the disparity between the lack of LTR-LTR divergence in individual *Junos* and the substantial values of Ks between copies requires explanation. The simplest possibility is that the most divergent and the least divergent *Juno* copies underwent at least two independent entries from an unidentified donor or donors. Less plausibly, the LTRs of individual elements may have been homogenized by conversion, observed in LTR retrotransposons only rarely (*e.g.* Johnson and Coffin, 1999). If so, such conversion would have had to be confined to LTRs belonging to the same element and not extending into adjacent regions on either side.

3.4. Phylogenetic placement.

Of special interest is the place occupied by *Vesta* and *Juno* in the phylogeny of LTR retrotransposons. Initial phylogenetic analysis grouped the *Vesta* RT-RH and IN domains with chromoviruses (not shown), although examination of the IN region showed the presence of the GPF/Y domain and the lack of a chromodomain. We then extended

the alignment to include the entire *pol* region (PR-RT-RH-IN). The tree was rooted with vertebrate retroviruses, which are a sister lineage to Metaviridae (Malik and Eickbush, 1999, 2001). The retrotransposon RNase H domain was aligned with the tether (connection) domain of retroviruses, previously shown to represent a remnant of an ancestral retrotransposon-like RNase H domain (Malik and Eickbush, 2001). The retroviral sequences included in the alignment also contain the GPY/F domain, which was subsequently lost in most retroviruses. Thus, the sequences employed in the alignment include all of the functional domains present in the *pol* region.

Although the two bdelloid LTR retrotransposons identified in this study are quite different from each other (25% and 50% aa identity of the *gag* and *pol* regions, respectively), they were found to form a distinct clade, with a bootstrap support value of 95% in neighbor-joining analysis, and 100% clade credibility value in Bayesian analysis (Fig. 5). The *DGLT-A* LTR retrotransposon from the social amoeba *Dictyostelium discoideum* (Glockner *et al*., 2001) appears to be most closely related to *Vesta* and *Juno*, although it cannot be placed in the same clade with comparable confidence (the bootstrap support in a neighbor-joining phylogeny is 46%). *DGLT-A* contains a single ORF, and none of the neighboring clades carry ORF3, suggesting its independent acquisition by *Vesta* and *Juno*. The bdelloid elements appear to form a sister clade to chromoviruses (Marin and Llorens, 2000; Kordis, 2005), a large and diverse clade of LTR retrotransposons with wide phylogenetic distribution, for which the major synapomorphy is the acquisition of the chromodomain C-terminally to the GPF/Y domain of the integrase.

3.5. Genomic environment.

The sequenced *Vesta*-containing cosmid clearly derives from a subtelomeric location, as seen in its localization by FISH and also in the presence of interspersed telomeric

repeats and *Athena* retroelements, as well as various DNA transposons (Arkhipova *et al*., 2003; E.G. and I.A., submitted). One of the *Juno* copies (*Juno4*) is also located on a sequenced telomeric fosmid, as judged by the presence of a long stretch of *A. vaga* telomeric repeats at one end of the fosmid, followed by two *Athena* retroelements. This copy, the only one to carry a 1-bp difference between left and right LTRs, is slightly 3' truncated (at nt 167 of the 3' LTR, just after the TATATA motif) and joined to a 3'-truncated *Athena*, indicating either a deletion of the intervening region, or gene conversion between the 3'-truncated *Juno4* and another telomere. The fosmid containing *Juno5*, initially selected by hybridization to the *A. vaga* telomeric repeat probe, contains a hAT-like transposase and could also be located in a subtelomeric region. In addition, a *gag* fragment was found on a *P. roseola* telomere adjacent to telomeric repeats in our telomere cloning experiments (E.G. and I.A., submitted).

Given the low copy number, such high incidence of occurrence in telomeric regions appears non-random. Any insertion site specificity at the level of nucleotide sequence, however, may be ruled out, since the TSDs and the immediate 0.6-0.7 kb flanking regions of the remaining *Juno* copies exhibit no sequence similarity, although the possibility of regional insertion specificity, broadly targeted towards specific chromatin domains, remains open. The sequenced flanking regions of the remaining *Junos* reveal no matches to known genes, with the exception of the most divergent copy, *Juno1* (96-97% identity to other four copies). It is inserted into an apparent ORF with homology to ligand-gated ion channel receptors, a multigene family containing 30-60 members in most eukaryotic species (Chiu *et al*., 2002; Iwama and Gojobori, 2002). Overall, as contrasted with the high density of structural genes seen in more proximal regions of bdelloid genomes (D. Mark Welch, J. Mark Welch and M. Meselson, unpublished), their lack in sequenced regions surrounding *Vesta* and *Juno* argues in favor of their

enrichment in subtelomeric heterochromatin, an environment highly enriched in other transposable elements as well, including *Athena* retroelements and DNA transposons (Arkhipova and Meselson 2005b; E.G. and I.A., submitted).

4. Discussion

In this study, we identified and characterized two LTR retrotransposons in bdelloid rotifers of two species, *P. roseola* and *A. vaga*, belonging to different families that separated tens of millions of years ago. Both retrotransposons exhibit typical structural features of retroviruses, including the presence of the third ORF corresponding to an *env*-like region. The *env*-like genes, with their potential to mediate horizontal entry into host cells, have been acquired by LTR retrotransposons at least six times, in three cases from a viral source (Malik *et al*., 2000). The well-studied insect errantiviruses from the gypsy clade were shown to have acquired their *env* region from baculoviruses, and the *env* regions of the nematode *Cer* and *Tas* retroviruses were acquired from a phlebovirus and a herpesvirus, respectively. So far, only one virus from aquatic invertebrates has been sequenced, a large DNA virus that can pass between rotifer and shrimp hosts (Yan *et al*., 2004). Although its ORFs do not reveal similarities with bdelloid *env*-like genes, additional viruses inhabiting aquatic invertebrates will hopefully be characterized in the future (*e.g*. Comps *et al*., 1991).

The simplest explanation of our findings, consistent with the presence of an *env*-like region in both elements, is that *Vesta* and *Juno* entered their respective bdelloid hosts as retroviruses from unidentified donors sometime after *P. roseola* and *A. vaga* separated. The identity of the left and right LTRs in *Vesta* and in four out of five *Junos* indicates that all of them inserted relatively recently, consistent with our failure to find decayed copies. In contrast, the synonymous site divergence between different copies of

Juno, ranging up to 5% or more, suggests that they have been diverging for a much longer time, apparently in their donor host(s), before entering *A. vaga*. Consistent with the Ka/Ks values, purifying selection would have operated on ORFs 1 and 2 during retrotransposition in their previous host, in the production of infective viruses, and during whatever limited transposition may have occurred within *A. vaga*. Being analogous to antigenic sites, the *env* coding regions may have been under both purifying and positive selection as in retroviruses, resulting in Ka/Ks values near unity.

The low copy numbers of *Vesta* and *Juno* (Fig. 3) and their solo LTRs (data not shown) may have several explanations, including relatively recent arrival in their present hosts. But our failure so far to discover other LTR retrotransposons suggests that additional factors may be at work, tending to deplete bdelloid genomes of retrotransposons generally. One possibility is suggested by the recent discovery that bdelloid rotifers are highly resistant to ionizing radiation (E.G. and M.M., unpublished) and the implication that such resistance is an adaptation to repair high levels of DNA breakage experienced during the desiccation and rehydration to which bdelloids are frequently exposed in their ephemerally aquatic habitats. Such repair may favor ectopic crossing-over of repeated elements such as transposons, with selection then acting against inviable translocations. Although *Vesta* and *Juno*, like other retroviruses and LTR retrotransposons, may have insertion site preferences determined by tethering to specific chromosomal proteins (Bushman, 2003), *e.g.* similar to yeast Ty5 (Xie *et al.*, 2001), ectopic crossing-over could also account for the preferential telomeric localization of these elements and of bdelloid DNA transposons (Arkhipova and Meselson, 2005b), as rearrangements confined to regions lacking essential genes proximal to the crossover breakpoints would presumably not be detrimental.

Acknowledgements

We thank the US National Science Foundation (MCB-0614142) and the National Institutes of Health (5R01GM072708-02) for financial support.

References

Arkhipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G., Ilyin, Y.V., 1986. The steps of reverse transcription of Drosophila mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. Cell 44, 555-563.

Arkhipova, I.R., Lyubomirskaya, N.V., Ilyin, Y.V., 1995. Drosophila retrotransposons. Austin, TX: RG Landes Co.

Arkhipova, I.R., 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. Genetics 139, 1359-1369.

Arkhipova, I., Meselson, M., 2000. Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci USA 97, 14473-14477.

Arkhipova, I., Meselson, M., 2005a. Deleterious transposable elements and the extinction of asexuals. Bioessays 27, 76-85.

Arkhipova, I.R., Meselson, M., 2005b. Diverse DNA transposons in rotifers of the class Bdelloidea. Proc Natl Acad Sci USA. 102, 11781-11786.

Arkhipova, I.R., Pyatkov, K.I., Meselson, M., Evgen'ev, M.B., 2003. Retroelements containing introns in diverse invertebrate taxa. Nat Genet. 33, 123-124.

Bae, Y.A., Moon, S.Y., Kong, Y., Cho, S.Y., Rhyu, M.G., 2001. CsRn1, a novel active retrotransposon in a parasitic trematode, Clonorchis sinensis, discloses a new phylogenetic clade of Ty3/gypsy-like LTR retrotransposons. Mol Biol Evol. 18: 1474-1483.

Barton, N.H., Charlesworth, B., 1998. Why sex and recombination? Science 281, 1986-1990.

Bell, G., 1982. The Masterpiece of Nature: The Evolution and Genetics of Sexuality. Berkeley: University of California Press.

Boeke, J.D., Eickbush, T.H., Sandmeyer, S.B., Voytas, D.F., 1999. In: Virus Taxonomy: VIIth Report of the International Committee on Taxonomy of Viruses, F.A. Murphy, ed. New York: Springer-Verlag.

Boeke, JD., Eickbush, T., Sandmeyer, SB., Voytas, DF., 2004. Metaviridae. In: Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses, C.M. Fauquet, ed. New York: Academic Press.

Bushman, F., 2003. Targeting survival: Integration site selection by retroviruses and LTR retrotransposons. Cell 115, 135-138.

Chiu, J.C., Brenner, E.D, DeSalle, R., Nitabach, M.N., Holmes, T.C., Coruzzi, G.M., 2002. Phylogenetic and expression analysis of the glutamate-receptor-like gene family in Arabidopsis thaliana. Mol Biol Evol. 19, 1066-1082.

Coffin, J.M., Hughes, S.H., Varmus, H.E., 1997. Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Comps, M., Mari, J., Poisson, F., Bonami, J.R., 1991. Biophysical and biochemical properties of an unusual birnavirus pathogenic for rotifers. J Gen Virol. 72, 1229-1236.

Friesen, P.D., Nissen, M.S., 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol Cell Biol. 10, 3067-3077.

Glockner, G., Szafranski, K., Winckler, T., Dingermann, T., Quail, M.A., Cox, E., Eichinger, L., Noegel, A.A., Rosenthal, A., 2001 The complex repeats of *Dictyostelium discoideum*. Genome Res. 11, 585-594.

Goodwin, T.J., Poulter, R.T., 2002. A group of deuterostome Ty3/ gypsy-like retrotransposons with Ty1/ copia-like pol-domain orders. Mol Genet Genomics 267, 481-491.

Heyman, T., Agoutin, B., Friant, S., Wilhelm, F.X., Wilhelm, M.L., 1995. Plus-strand DNA synthesis of the yeast retrotransposon Ty1 is initiated at two sites, PPT1 next to the 3' LTR and PPT2 within the pol gene. PPT1 is sufficient for Ty1 transposition. J Mol Biol. 253, 291-303.

Iwama, H., Gojobori, T., 2002. Identification of neurotransmitter receptor genes under significantly relaxed selective constraint by orthologous gene comparisons between humans and rodents. Mol. Biol. Evol. 19, 1891–1901.

Johnson, W.E., Coffin, J.M., 1999. Constructing primate phylogenies from ancient retrovirus sequences. Proc Natl Acad Sci USA 96:10254-102560.

Kondrashov, A.S., 1993. Classification of hypotheses on the advantage of amphimixis. J Hered 84, 372–387.

Kordis, D., 2005. A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. Gene 347, 161-173.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5, 150-163.

Li, W.H., 1997. Molecular Evolution. Sinauer, Sunderland, MA.

Malik, H.S, Eickbush, T.H., 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J Virol. 73, 5186-5190.

Malik, H.S., Eickbush, T.H., 2001. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res. 11, 1187-1197.

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

Marin, I., Llorens, C., 2000. Ty3/Gypsy retrotransposons: description of new Arabidopsis thaliana elements and evolutionary perspectives derived from comparative genomic data. Mol. Biol. Evol. 17, 1040–1049.

Mark Welch, D., Meselson, M., 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science 288, 1211-1215.

Mark Welch, J.L., Mark Welch, D.B., Meselson, M., 2004. Cytogenetic evidence for asexual evolution of bdelloid rotifers. Proc Natl Acad Sci USA 101, 1618-1621.

Misseri, Y., Labesse, G., Bucheton, A., Terzian, C., 2003. Comparative sequence analysis and predictions for the envelope glycoproteins of insect endogenous retroviruses. Trends Microbiol. 11, 253-256.

Normark, B.B., Judson, O., Moran, N., 2003. Genomic signatures of ancient asexual lineages. Biol. J. Linn. Soc. 79, 69-84.

Otto, S.P., Lenormand, T., 2002. Resolving the paradox of sex and recombination. Nature Rev Genet 3:252–260.

Pamilo, P., Bianchi, N.O., 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol. 10, 271-281.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Smale, S.T., Baltimore, D., 1989. The "initiator" as a transcription control element. Cell 57, 103-113.

Spana, C., Harrison, D.A., Corces, V.G., 1988. The *Drosophila melanogaster* suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. Genes Dev. 2, 1414-1423.

Wilson, S., Matyunina, L.V., McDonald, J.F., 1998. An enhancer region within the copia untranslated leader contains binding sites for Drosophila regulatory proteins. Gene 209, 239-246.

Wright, D.A., Voytas, D.F., 2002. Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses. Genome Res. 12, 122-131.

Xie, W., Gai, X., Zhu, Y., Zappulla, D.C., Sternglanz, R., Voytas, D.F., 2001. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol. 21, 6606-6614.

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

Yan, D.C., Dong, S.L., Huang, J., Yu, X.M., Feng, M.Y., Liu, X.Y., 2004. White spot syndrome virus (WSSV) detected by PCR in rotifers and rotifer resting eggs from shrimp pond sediments. Dis Aquat Organ. 59, 69-73.

LEGENDS TO FIGURES

Figure 1. Structural organization of (**A**) *Vesta* and (**B**) *Juno* retrotransposons. Shaded triangles, LTRs; open boxes, ORFs. Functional domains within each ORF are discussed in the text. The predicted introns connect two putative exons in ORF3, although may not necessarily be functional in bdelloid hosts. Striped boxes denote the corresponding probes used in hybridization experiments.

Figure 2. Multiple sequence alignment of *Vesta* and *Juno* with known representatives of Metaviridae and Retroviridae. Shown are the most conserved regions in the *gag*-like ORF and in the PR, RT, RH, and IN domains of *pol.* Highly conserved residues are denoted by asterisks. Functionally important residues essential for binding or catalysis (listed in Xiong and Eickbush, 1990; Malik and Eickbush, 1999, 2001) are indicated on the top. The sequences from diverse clades, used in the alignment, including several uncharacterized retroelements, were retrieved as the top matches to the *Vesta* and *Juno* queries in a search of the CDD (conserved domain database, NCBI). The host species and GenBank identifier (gi) numbers of retroelements are as follows: *DGLT-A1* from *Dictyostelium discoideum* (11527878); *sushi* from *Fugu rubripes* (6425167); *Athila* from *Arabidopsis thaliana* (4417306); *Galadriel* from *Lycopersicon esculentum* (4235644); *Ty3* from *Saccharomyces cerevisiae* (173086); *Woot* from *Tribolium castaneum* (1213461); *blastopia* from *Drosophila melanogaster* (415797); *Cer1* from *Caenorhabditis elegans* (557717); *MuLV* from *Mus musculus* (535517); uncharacterized retroelements from *C. elegans* (K03D3.8, 3878215) *Oryza sativa* (13486715), *Glomerella cingulata* (10946131), *Zea mays* (2832244), and *Danio rerio* (68394879). The Gag alignment, in its CC-HC part, is a continuation of the sequences on the left, except for two cases: line

3 shows the second Zn knuckle of Juno, 7 aa downstream from the first one, and line 7 shows the retroviral Zn knuckle from MuLV.

Figure 3. Characterization of the putative env-like proteins encoded by (**A**) *Vesta* and (**B**) *Juno*. The putative N-linked glycosylation sites (NxS/T) are shaded; the potential furin-like protease cleavage sites (RxxR) and the canonical fusion tripeptide (FxG) motifs (Misseri *et al*., 2003) are in bold italics. The transmembrane (TM) regions predicted by PSORT are underlined. Two TM regions are found in *Juno2* (shown in the figure); only one region appears in *Juno3, 5*, and *4*; the most divergent *Juno1* lacks detectable TM regions. Cysteine residues are in boldface. Predicted intron locations are shown by triangles. We have not determined the exact structure of subgenomic RNAs used for expression of the *env* regions, thus the first methionine in these ORFs does not necessarily correspond to the actual N-terminus of env-like proteins, normally expected to have a signal peptide. Below the amino acid sequence of each ORF is its graphical representation (shaded boxes, putative transmembrane domains; vertical arrows, potential host protease cleavage sites; Y, putative N-glycosylation sites) and the corresponding Kyte-Doolittle hydrophobicity plot generated by ConPred2, with predicted TM segments shown as shaded boxes (http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/).

Figure 4. Southern blot hybridization estimates of copy numbers of *Vesta* (**A**) and *Juno* (**B**) in genomic DNA. Hybridization probes are indicated in Fig. 1. (A) *P. roseola* genomic DNA digested with (1) *Eco*R1, (2) *Hin*dIII and (3) *Pvu*II. *Eco*R1 has a single recognition site within *Vesta* and yields no internal *Vesta* fragments; *Hind*III yields a *ca*. 5-kb internal fragment, in addition to fragments extending into flanking DNA; and *Pvu*II has no internal recognition sites. (B) *Pvu*II digests of genomic DNA from *A. vaga* (1) and a related

species *A. ricciae* (2), which yields weak hybridization. A *ca*. 4-kb internal *Juno* fragment is generated upon *Pvu*II digestion, in addition to fragments extending into flanking DNA.

Figure 5. Phylogenetic placement of *Vesta* and *Juno*. The phylogram shows the results of the neighbor-joining analysis, with the clade support values from 1000 bootstrap replications indicated below the branches, and with the clade credibility values obtained from Bayesian analysis indicated above the branches, if any of the values exceeds 50%. Previously known clades of Metaviridae (Malik and Eickbush, 1999; Bae *et al*., 2001; Kordis, 2005) are indicated on the right. The Gmr clade, with the inverse IN-RT order (Goodwin and Poulter, 2002), was not included in the alignment. All clades are significantly supported, but in most cases the branching order of the clades, as in previous studies, cannot be resolved. The elements *Woot* and *Cigr1* apparently do not belong to known clades, and their placement differs in neighbor-joining and Bayesian analyses. Amino acid sequences are from the datasets in the above references, which were used to define these clades; also shown are several sequences listed in the legend to Fig. 2.

Table 1. Characteristics of LTRs and adjacent regions in *Vesta* and *Juno*.

	TSD	LTR	TCAGT	AATAAA	LTR ends	PBS	PPT
Vesta	GTGGA	343 bp	223	245	TGTTA...TAACA	TTTTGGTGGAGAAGTTGG	CAAACGGTGAAGG
Juno	GTTTC, ATATA, ACATT, ATCAT	283 bp	186	245	TGTTA...TAATA	TTTGGTGCCGTGACCGG	GCGTTGGGTCGAC

* The numbers for Juno correspond to the consensus sequence of three copies, *Juno2,3* and *5*, which are 99.7% identical. TCAGT, position of the predicted RNA start site; AATAAA, position of the polyadenylation signal within the LTR. The TGG complementary to the 3' CCA of one of the host tRNAs is underlined. PBS, primer-binding site; PPT, presumptive polypurine tract; TSD, target-site duplication (for *Juno1, 2, 3*, and *5*, in that order).

Table 2. Comparison of the *Vesta* and *Juno* coding regions.

	LTR-ORF1	ORF1	ORF1-2	ORF2*	ORF2-3 HVR	ORF3*	ORF3-LTR*
Vesta	242 bp	601 aa	-1 f/s, 34bp overlap	1269 aa	210 bp	299 (440) aa	350 (0) bp
Juno	294 bp	327 aa	-1 f/s, 55bp overlap	1302 aa	550-575 bp	520 (597) aa	254 (0) bp

* The length of ORF2 and ORF3 is calculated from the first ATG codon, although the actual proteins may slightly differ in length because of proteolytic processing or variation in splicing patterns. HVR, hypervariable region between ORF2 and ORF3 in different *Juno* copies, likely corresponds to an intron, by analogy to retroviruses, which normally express their *env* genes from a subgenomic RNA generated by splicing. The numbers in parentheses indicate lengths after introduction of an additional intron in order to generate an extended ORF3 (see text).

Table 3. Divergence between five genomic *Juno* copies (*Juno1-5*) in the coding regions.

	ORF1					ORF2					ORF3				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		.275	.247	.355	.249		.400	.400	.444	.392		.852	.852	.733	.853
2	.060		n.a.	.479	n.a.	.036		n.a.	.544	.207	.020		n.a.	1.325	n.a.
3	.067	.006		.393	.260	.036	.000		.588	.620	.020	.000		1.325	n.a.
4	.053	.029	.035		.424	.038	.014	.014		.510	.014	.009	.009		1.327
5	.060	.000	.006	.029		.037	.001	.001	.015		.020	.000	.000	.009	

For each ORF, the Ks values (bottom left) and Ka/Ks ratios (top right) calculated by the DIVERGE program are presented. n.a., ratio could not be calculated (Ks or Ka equals zero). Values for ORF3 include exon 1 only; the putative exon 2 contains no nucleotide substitutions, but includes two homopolymeric stretches in which polymorphisms were apparently introduced by replication slippage, and was not included in calculations.

Figure 1

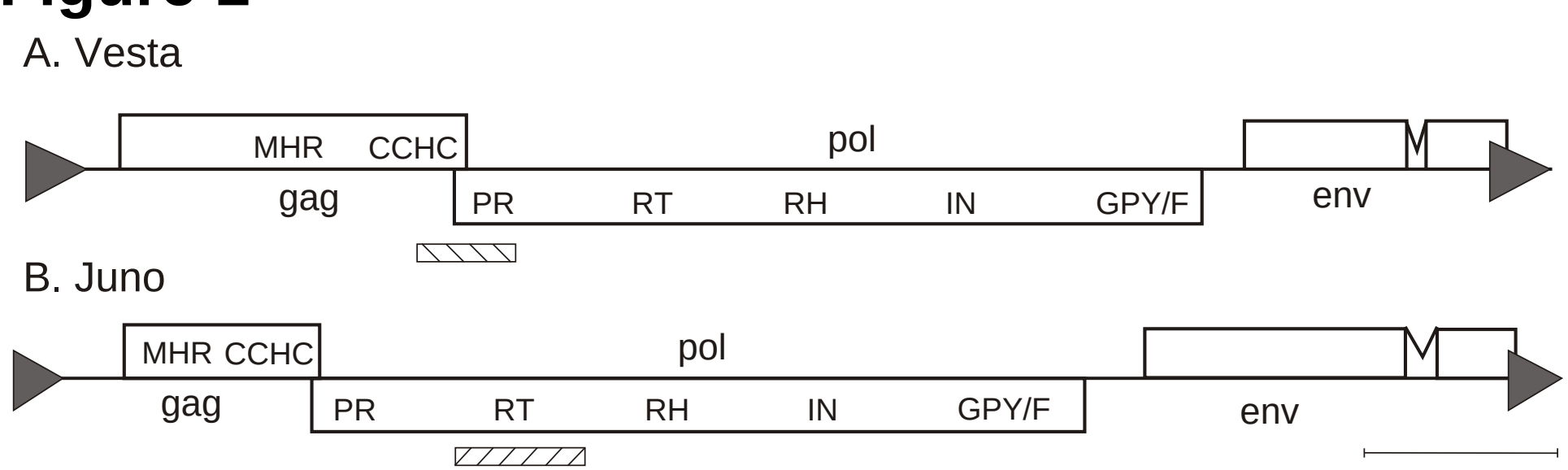
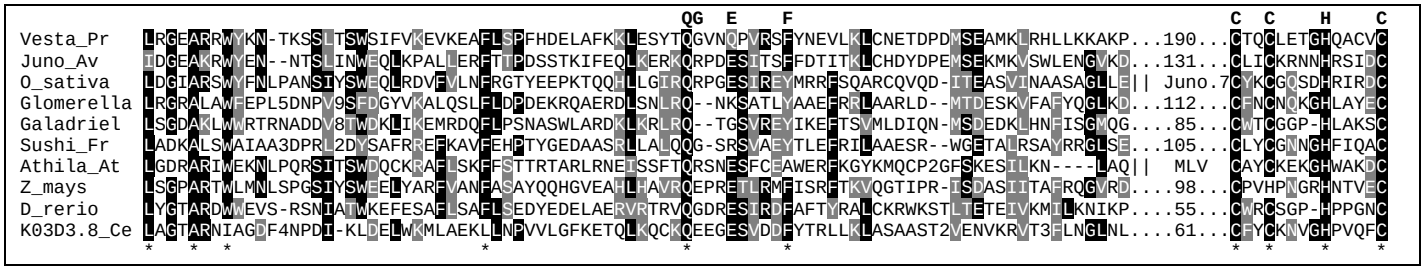
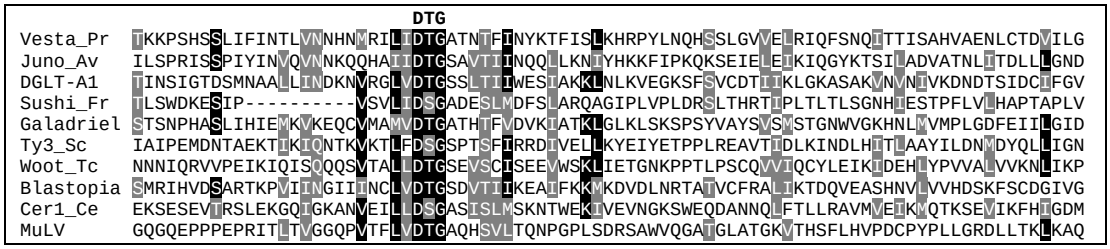

A. Vesta
MHR
CCHC
gag
pol
PR
RT
RH
IN
GPY/F
env
B. Juno
MHR CCHC
gag
pol
PR
RT
RH
IN
GPY/F
env

Figure 2

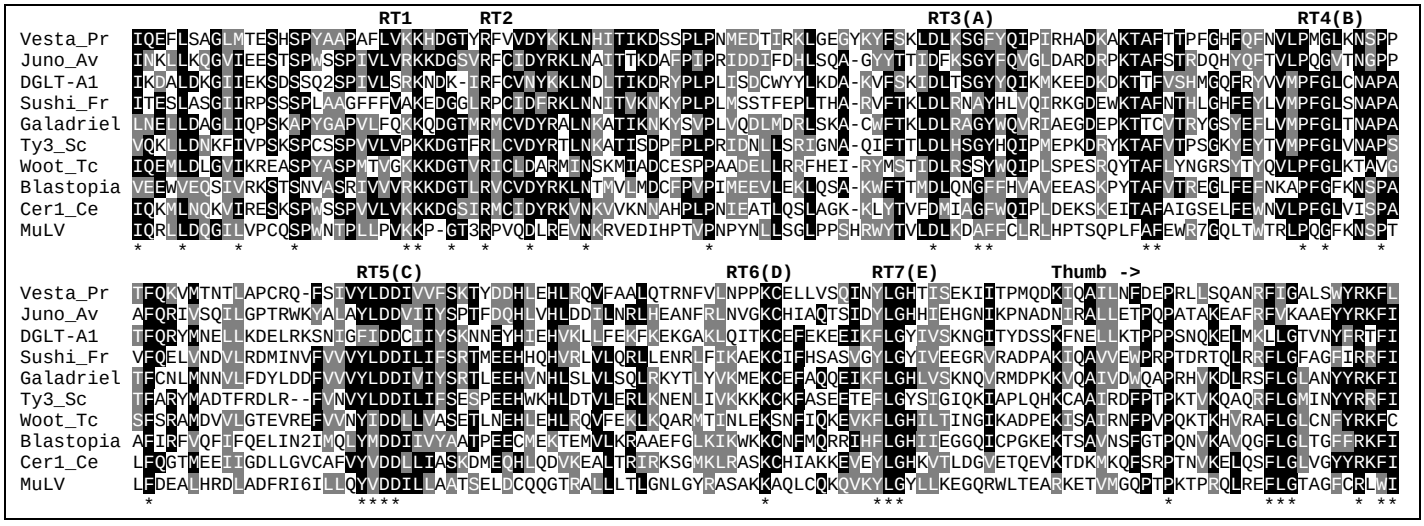
GAG



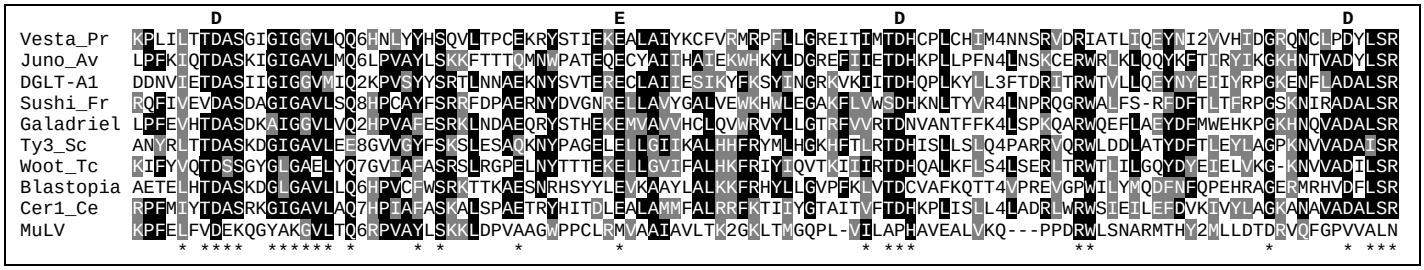
PR



RT



RH



IN

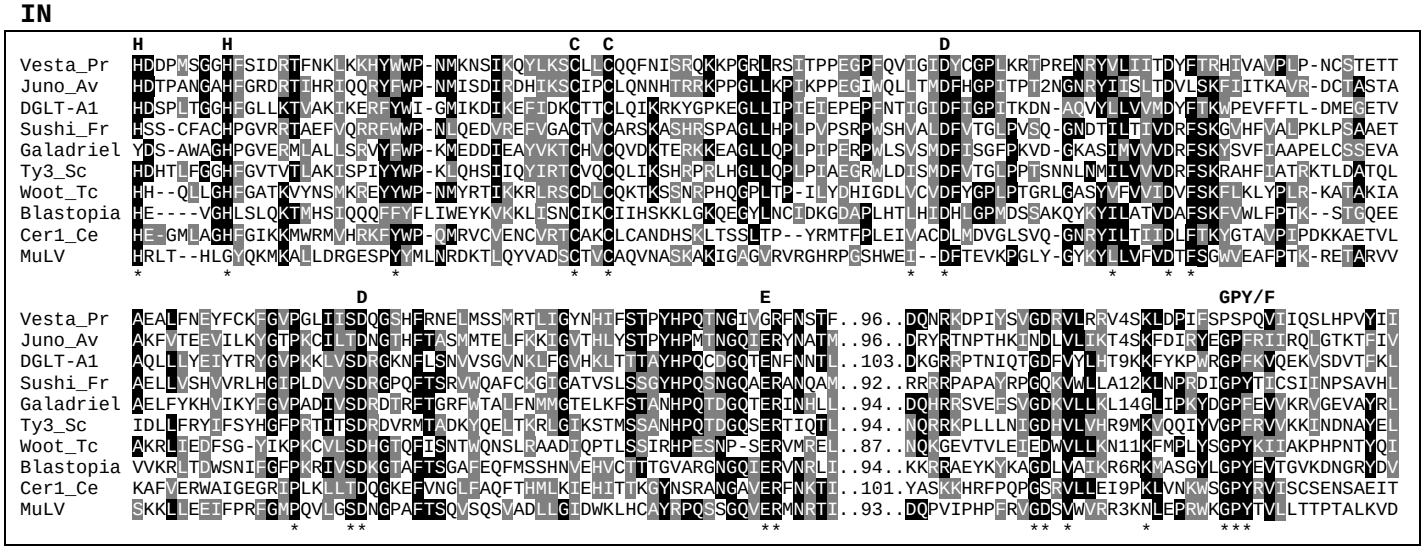
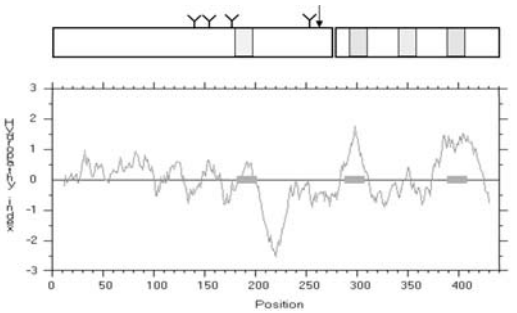


Figure 3

A. *Vesta*

MNLPALPNVPNYFVLTDSHGKFVPPFITTPTYSIHIMAIFGLKWINSYNP
AL**C**TATLLSTSHLHQLLSSANSIMFLIGTNSLR**C**TPAVTVITQVSDLIHL
LRFQHPHLSNKHSITLVA**C**FP**C**VKPIYPLNTTHSLSNNITQYNTMLMDLS
AAMNFTVIDFHVLEHHLGFDRMHLAHQHKHLVHLSIINYFTYLSSIPTPS
TVATIGRSAEAKAR***RNQR***RHHKQAIKQRQHFLTRTIHSSWSVLSVKDYLH
KNDIKFVKLPPIHRNTLRIQFNNPFDLQVADANLSQN▲LTPFLFVVYLSS
LVSFFFFAVEPKRYVSSPTGYAGRRIMYLT**C**SNDEHYRLQLSLFHYLLSS
DYGFHFHADII**C**RQESMKIHIDRRQTLIFKFVYLLFFSF**C**QTVKVLASTV
VTYVVANVLLFFSSL**C**MNW**C**SLTWTSVHTSFSTPPPDSKTR



B. *Juno*

MDRKSS**C**SRKYYNDRYVPPLEELLRRNERWNPKKAKKERELQEL**C**RQEKL
KPFEQPKHYSRYYIHEKTDMETMDKLIDEAKLTFNYTLDTEGDAYNRISA
ATIQVEFIRPNSPSIVIIIEVNYLPPISSPLFIKIRQL**C**SIIFTSNNRIY
AWGYVADELKSFMNLNLFSGQIKADGVNVQNEYDRYELYGLQAVVKSTFG
QYLDKTATLAKWS**C**GIDRSLDTYKPKYVHGREYDYRVQEETKYRHMLEEY
AINDVFAVTKVAYDMNLIKFVLTPPATVENEKDVPVESTTQQEPTVELKP
PNRDEPEAHAYIDSREEEHPGQQESISIELELTPSDLDIGIFDHENESSP
VEMEEEPIESPSQKVQPTPYVYEPISDDEFPEVMKLHRPFPQPDPTPEER
KRIHAQDESQKVRNSYYLGPQLYLQENPTPNQISNRRR***RSNR***YRSEVIYP
VHRLFKHAHIKQILRSKNIQFLNINIKHGKV***FIG***▲TYIYFFPFLPSFEMY
TRVSFLFSLFFSFFLYEPYIYIYTYRHFSFVSSFVFSLFLSRSLLIFFFR
SVRYM**C**ADTHRSALGRLLQATPISSYSSLLFLFIDLVFHLSLDLDPFF

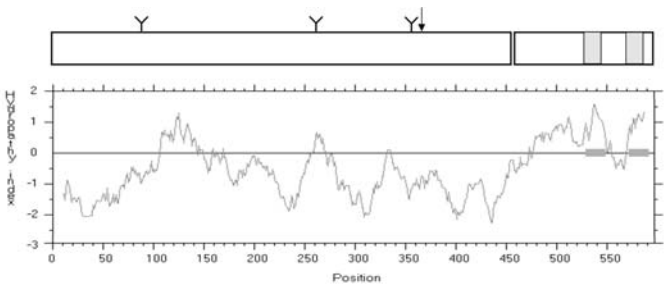


Figure 4

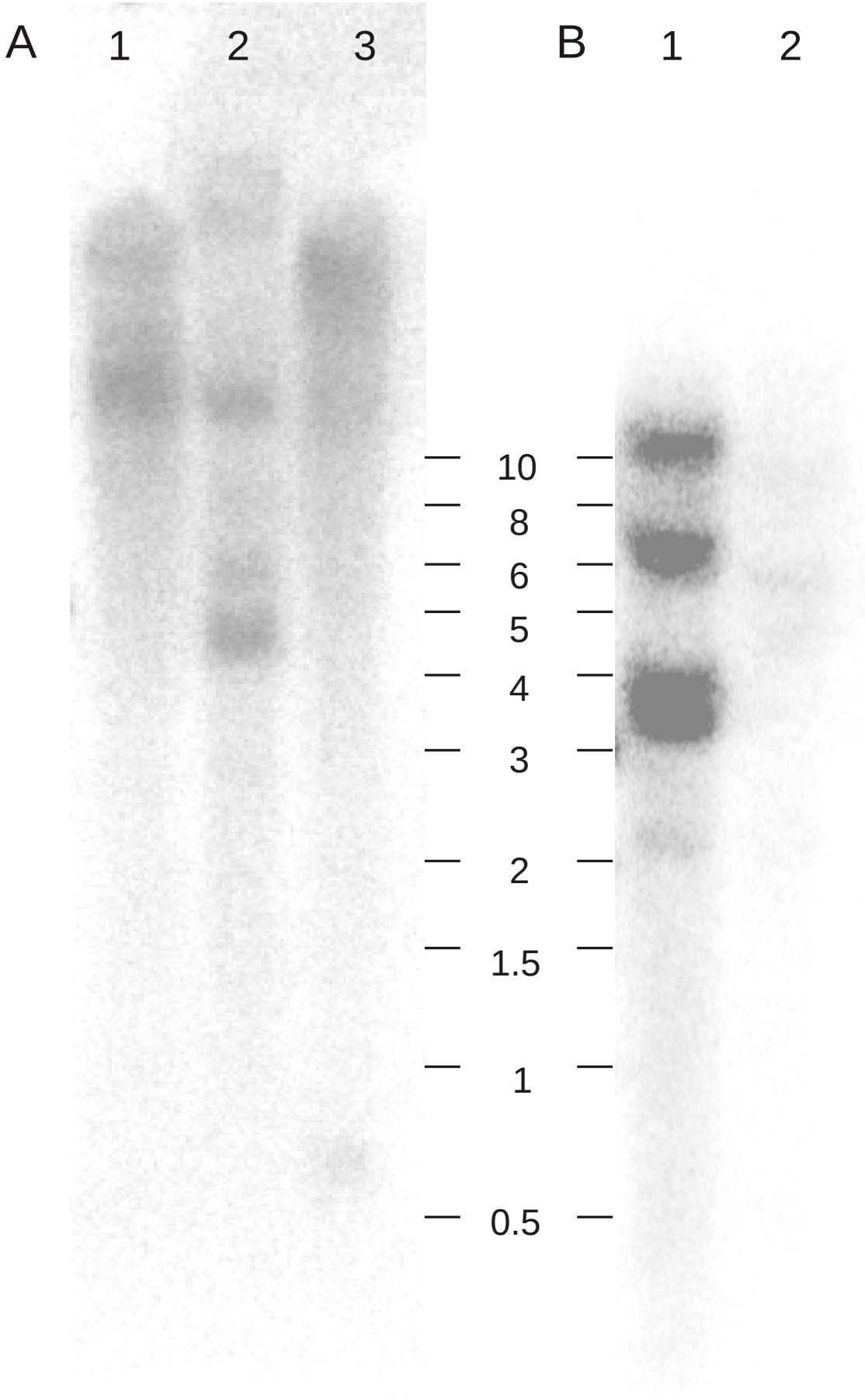
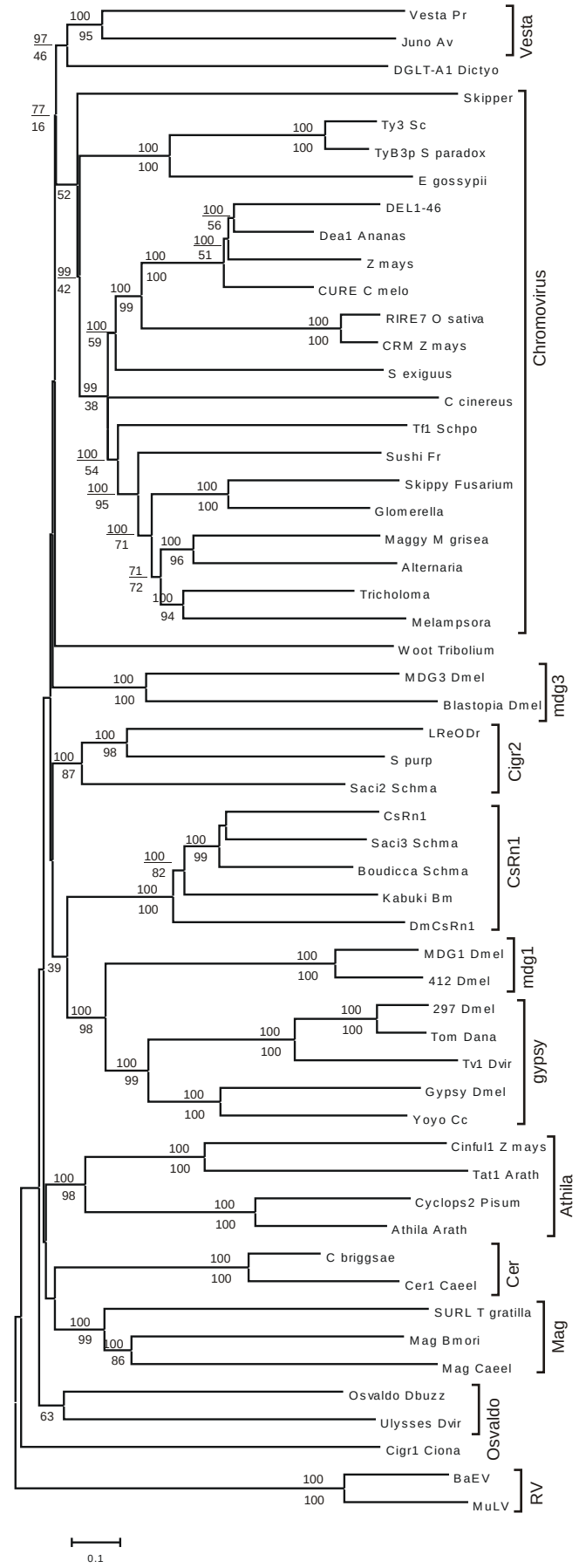


Figure 5


Vesta Pr
Juno Av
DGLT-A1 Dictyo
Vesta
Skipper
Ty3 Sc
TyB3p S paradox
E gossypii
DEL1-46
Dea1 Ananas
Z mays
CURE C melo
RIRE7 O sativa
CRM Z mays
S exiguus
C cinereus
Tf1 Schpo
Sushi Fr
Skippy Fusarium
Glomerella
Maggy M grisea
Alternaria
Tricholoma
Melampsora
Chromovirus
Woot Tribolium
MDG3 Dmel
Blastopia Dmel
mdg3
LReODr
S purp
Saci2 Schma
Cigr2
CsRn1
Saci3 Schma
Boudicca Schma
Kabuki Bm
DmCsRn1
MDG1 Dmel
412 Dmel
mdg1
297 Dmel
Tom Dana
Tv1 Dvir
Gypsy Dmel
Yoyo Cc
gypsy
Cinful1 Z mays
Tat1 Arath
Cyclops2 Pisum
Athila Arath
Athila
C briggsae
Cer1 Caeel
Cer
SURL T gratilla
Mag Bmori
Mag Caeel
Mag
Osvaldo Dbuzz
Ulysses Dvir
Osvaldo
Cigr1 Ciona
BaEV
MuLV
RV
0.1