

Transcriptome Analysis of ESTs from a Chaetognath Reveals a Deep-Branching Clade of Retrovirus-Like Retrotransposons

Roxane M. Barthélémy, Jean-Paul Casanova and Eric Faure*

LATP, CNRS-UMR 6632, Evolution biologique et modélisation, case 5, Université de Provence, Place Victor Hugo, 13331 Marseille cedex 3, France

Abstract: Chaetognaths constitute a small marine phylum exhibiting several characteristic which are highly unusual in animal genomes, including two classes of both rRNA and protein ribosomal genes. As in this phylum presence of retrovirus-like elements has never been documented, analysis of a published expressed sequence tag (EST) collection of the chaetognath *Spadella cephaloptera* has been made. Twelve sequences representing transcript sections of reverse transcriptase domain of active retrotransposons were isolated from ~11,000 ESTs. Five of them are originated from *Gypsy* retrovirus-like elements, whereas the other are transcripts from a *Bel-Pao* LTR-retrotransposon, a *Penelope*-like element and LINE retrotransposons. Moreover, a part of a putative integrase has also been found. Phylogenetic analyses suggest a deep-branching clade of the retrovirus-like elements, which is in agreement with the probably Cambrian origin of the phylum. Moreover, retrotransposons have not been found in telomeric-like transcripts which are probably constituted by both vertebrate and arthropod canonical repeats.

Keywords: Retrovirus-like, retrotransposons, Chaetognatha, *Spadella cephaloptera*, reverse transcriptase.

INTRODUCTION

Chaetognaths are a small marine phylum living in various habitats, but most of them are among the most abundant planktonic organisms [1]. Their body is constituted of three parts, the head, trunk and tail, separated by septa [2]. These animals are protandric hermaphrodites; the ovaries lie in the trunk on both sides of the gut, while the testis are in the tail (reviewed in [3]). Their phylogenetic position remains enigmatic, although recent molecular analyses suggest a protostome affinity [4-7]. Casanova *et al.* [8] showed that chaetognaths can be considered as a model animal.

The chaetognath genomes exhibit several molecular singularities including paralog of both ribosomal RNA genes and ribosomal protein genes [9-11] and conservation of extremely divergent paralogous sequences suggesting a low rate of gene conversion [12]. Moreover, *in situ* hybridizations have shown that each type of both 18S and 28S rRNA paralogs is important for specific cellular functions [13-15]. The causes of these features are unknown, even if an allopolyploid event has been suggested [12]. As it is well known that mobile genetic elements (MGEs, also called transposable elements) can strongly impact genome evolution, knowledge of these elements about few is known in chaetognath [16] could be a fruitful contribution.

MGEs are ubiquitous in a wide range of living organisms; however, they make up a large fraction of genome sizes which is evident through the C-values of only pluricellular eukaryotes. These elements, which can transpose from one location to another within the genome, are known to be

one of the causes of large scale genome reorganization [17]. Although regarded as a selfish DNA with negative impact on the host, MGEs have been shown to contribute significantly to gene evolution [18]. Now these elements are regarded as one of the principal forces driving the evolution of eukaryotic genomes [19,20]. Due to the great number of known MGEs (several thousands) and as new types of mobile repeats are discovered at a rapid rate, a unified classification system for eukaryotic transposable elements has recently been proposed, designed on the basis of the transposition mechanism, sequence similarities and structural relationships [21]. MGEs are divided into two classes. Class I retrotransposons replicate *via* an RNA intermediate; the key enzyme of this mechanism is the reverse transcriptase (RT), each complete replication cycle produces one new copy. Class II transposons, which are out of our topic, move as a DNA segment by a classical “cut-and-paste” mechanism.

Retrotransposons have been divided into five orders on the basis of their mechanistic features, organization and reverse transcriptase (RT) phylogeny: LTR-retrotransposons, *DIRS*-like elements, *Penelope*-like elements, LINES and SINEs. The LTR-retrotransposons are retrovirus-like elements containing long terminal repeat (LTR) and ORFs for at least *gag*, a structural protein for virus-like particles, and *pol*. *Pol* encodes an aspartic proteinase (PR), reverse transcriptase, RNase H, and a DDE integrase (IN). LTR retrotransposons also contain specific signals for packaging, dimerization, reverse transcription and integration. The two main superfamilies *Gypsy* and *Copia*, differing in the relative order of integrase and RT domains. According to Wicker *et al.* [21], the other members are retrovirus, endogenous retroviruses (ERVs) and the *Bel-Pao* subfamily containing MGEs structurally similar to *Gypsy* or *Copia* elements but exhibit differences on RT phylogenies. Evolutionarily, LTR retrotransposons are closely related to retroviruses. Retroviruses have

*Address correspondence to this author at the LATP, CNRS-UMR 6632, Evolution biologique et modélisation, case 5, Université de Provence, Place Victor Hugo, 13331 Marseille cedex 3, France; Tel: 00 (33) 491 10 61 77; Fax: 00 (33) 491 10 62 25; E-mail: Eric.Faure@univ-provence.fr

a viral lifestyle through acquisition of an envelope protein added to various regulatory proteins. Retroviruses are distributed widely among vertebrates and may also occur in some invertebrates, for example, some members of the *Gypsy* family found in dipteran insects are able to infect new individuals [22]. Retrovirus can also be transformed into an LTR retrotransposon-like through inactivation or deletion of the domains that enable extracellular mobility and can only be inherited by vertical transmission through the germ line, this is the case of the endogenous retroviruses even if some of them can always be transmitted horizontally.

Penelope-like elements (PLEs) represent a new order of retroelements identified in more than 80 species including unicellular animals, fungi and plants [23]. These elements code for a protein that represents a fusion between a reverse transcriptase and a GIY-YIG endonuclease. They encode an RT that is more closely related to telomerase than to the RT from LTR retrotransposons or LINES; moreover, members of this order have also LTR-like sequences that can be in a direct or an inverse orientation. The principal other retrotransposon orders are the LINES and the SINES. The LINES lack LTRs, can reach several kilobases in length, and are found in all eukaryotic kingdoms. Autonomous LINES encode at least an RT and a nuclease in their *pol* ORF for transposition, they often display a poly(A) tail at their 3' end. The SINE elements are non-autonomous elements and as they do not contain RT gene, they rely on the activity of RT proteins encoded by LINES to retrotranspose [24]. They originated from accidental retrotransposition of various polymerase III transcripts and possess an internal III promoter, allowing them to be expressed. SINES belong to retrosequences, a group containing all the sequences arisen by reverse transcription of ribosomal, messenger and small stable RNAs [25,26].

A previous study using degenerate primers gave positive results during screening of *Sagitta* sp. for LINE-like and *Gypsy*-like reverse transcriptases, and also for *Mariner*-like transposases; however, only two LINE elements have been sequenced [16]. As in this phylum retrovirus-like sequences have never been documented, analysis of a published expressed sequence tag (EST) collection [6] of the chaetognath *Spadella cephaloptera* has been made.

MATERIAL AND METHODS

EST Sequence Identifications

The chaetognath EST collection used has been annotated by Marlétaz *et al.* [6]. Since then, a great number of new sequences have been deposited in databases. As a great level of sequence diversity could be found even in the same retrotransposon subfamily, this has necessitated a new analysis has been made. Since amino acid sequences are more useful to detect homology over long periods, the EST sequences were translated in all six reading frames and compared to the sequences in the NCBI nr and Swissprot protein databases. Sequences that did not match were further compared against the GenBank and dbEST nucleotide databases (Blastn). Among 11,254 sequences, thirteen showed similarity (e -value $< 10^{-5}$) to previously described retroelement sequences.

Blast and Phylogenetic Analyses

For each amino-acid sequences deduced from chaetognath ESTs which are homologous to retrotransposon genes are

automatically searched for in the full length proteins from NCBI NR protein database. At this step, the Figenix platform has been used to automatically detect homologs based upon robust phylogenetic reconstruction [27]. When the number of homologs automatically detected is lower than 20, BLAST-based datasets were constructed using BLASTp queries against NCBI NR protein database. The Figenix platform has also been used for these phylogenetic reconstructions. The robustness of the tree has been tested by bootstrap analyses with 1000 resamplings. As for some analyses the number of homologous sequences was too low (< 8), only the alignments are given using the Clustal W program [28].

RESULTS

Blast and Phylogenetic Analyses

The cDNA library has been made from mRNAs isolated from various embryonic stages of *Spadella cephaloptera* (from 0 to 48 hours after hatching) [6]. The 5'-ends of 11,254 clones from this library have been sequenced and after annotation analyses the homology relations have been assigned to 2396 clones corresponding to the transcripts of 792 different genes. Similarly to Marlétaz *et al.*'s annotation [6], our re-analysis suggests that thirteen ESTs represent transcript sections of active retroelements (Table 1). Three EST sequences are strictly identical and one sequence is internal to another EST. In spite the fact that alignments with published retrotransposons suggest that some EST sequences could partially overlap between them, it has been impossible to assemble these ESTs into contig due to significant nucleotide differences in the overlapping regions, suggesting that these ESTs belong to different MGEs although very close phylogenetically.

Generally, in retrotransposons the *gag* and *pol* genes are intact; however some are interrupted by inframe stop-codons, but some of the non-autonomous LTR-retrotransposons can transpose *via* trans-activation by autonomous partners. Only three of the chaetognath ESTs (CR950075, CR950076 and CR950197) do not contain stop codons; moreover, two ESTs have a microsatellite sequence linked to a part of the *pol* gene (Table 1). In addition, in the chaetognath EST collection, only sequence homologies against Pol domain has been found; indeed, it is well known that the *pol* gene is the most conserved among the retrotransposons and to elucidate the phylogenetic relationships between these elements, *pol*-based trees have been constructed since Xiong and Eickbush [29]. Our analyses also shown that twelve out of thirteen of the ESTs encode part of the RT domain and generally the COOH-terminal amino acid region, whereas an EST sequence exhibits some homologies with a putative integrase (Table 1 and Fig. 1).

Using Figenix platform, six phylogenetic trees have been obtained (Fig. 1). Three analyses reveal that the deduced sequences of five ESTs belong to the *Gypsy/Ty3* superfamily (Fig. 1A, B and C). These retrovirus-like elements have been found in animals, fungi and plants [21] and sequences belonging to these three taxa are present in the three phylogenies. Moreover, several clades of retrovirus-like elements have been found for the three taxa. Although the bootstrap values are very low, all these analyses suggest that the partial chaetognath sequences group with homologous regions of animal *Gypsy/Ty3* elements. In two analyses, the chaetog-

Table 1. Chaetognath ESTs Containing Transcript Sections of Retrotransposon Domains

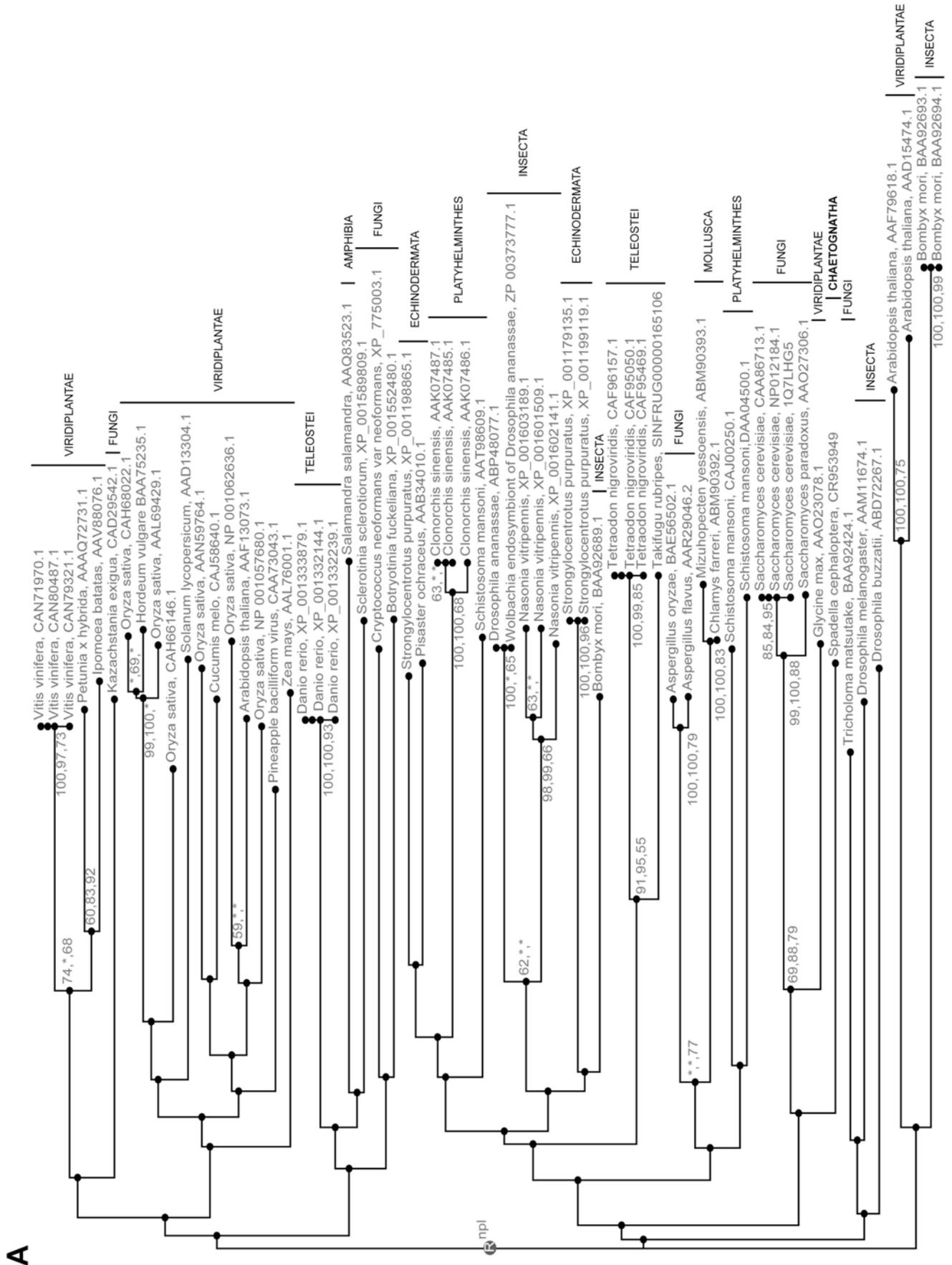
Type of Retrotransposon: Order and Superfamily for LTR- Retrotransposon	EST Acc. n°	Total Length - [Length of the Microsatellite Region]- (Length of the poly(A) Tail)	Data Concerning the Closer Homologous Complete Retrotransposon: Name of the Element (if Known), Taxon, Species, Acc. n°	Protein Domain Matching	Other Characteristic(s) of the ESTs
LTR-retrotransposon: Retrovirus-like, <i>Gypsy</i>	CR953949	641-(0)	<i>Tv1</i> , insect <i>Drosophila virilis</i> (AF056940)	RT-RNase H	One frameshift
LTR-retrotransposon: Retrovirus-like, <i>Gypsy</i>	CR953634 CR953418 CR953554	494-(23) 494-(23) 494-(23)	<i>Tv1</i> , insect <i>Drosophila virilis</i> (AF056940)	RT	The three sequences are strictly identical
LTR-retrotransposon: Retrovirus-like, <i>Gypsy</i>	CR952896	360-(0)	<i>Tv1</i> , insect <i>Drosophila virilis</i> (AF056940)	RT	
LTR-retrotransposon: <i>Bel-Pao</i>	CR950076 CR950075	784-(0) 682-(0)	<i>Kamikaze</i> , insect <i>Bombyx mori</i> [30]	RT	Shortest sequence internal to the longest
<i>Penelope</i> -like element (Ple)	CR950254	[53]-471-(26)	<i>Xena</i> , fish <i>Takifugu rubripes</i> (AAK58879)	RT	(CAA) _n microsatellite
<i>Penelope</i> -like element (Ple)	CR950255	706-[303]-(0)	<i>Xena</i> , fish <i>Takifugu rubripes</i> (AAK58879)	RT	(CAA) _n microsatellite
<i>Penelope</i> -like element (Ple)	CR941783	714-(30)	<i>Xena</i> , fish <i>Takifugu rubripes</i> (AAK58879)	RT	Complementary sequence
LINE	CR950197	574-(0)	<i>CRI</i> , testudine, <i>Platemys spixii</i> (AB005891)	RT	
LINE	CR950101	669-(0)	<i>CRI-1</i> , fish, <i>Danio rerio</i> (AB211149)	RT	
Putative integrase of an unknown element	CR950196	678-(30)	Urochordate <i>Oikopleura dioica</i> (AAS21408)	IN	Doubtful, homology with only one putative integrase

Several characteristics of each EST have given and of the closer homologous complete retrotransposon are reported. Abbreviations: (Acc. n°) Accession number, (RT) transcriptase inverse, (IN) integrase. The position of the number corresponding to the microsatellite sequence length reflects the position in the EST, i.e. in 5' or in 3' regions.

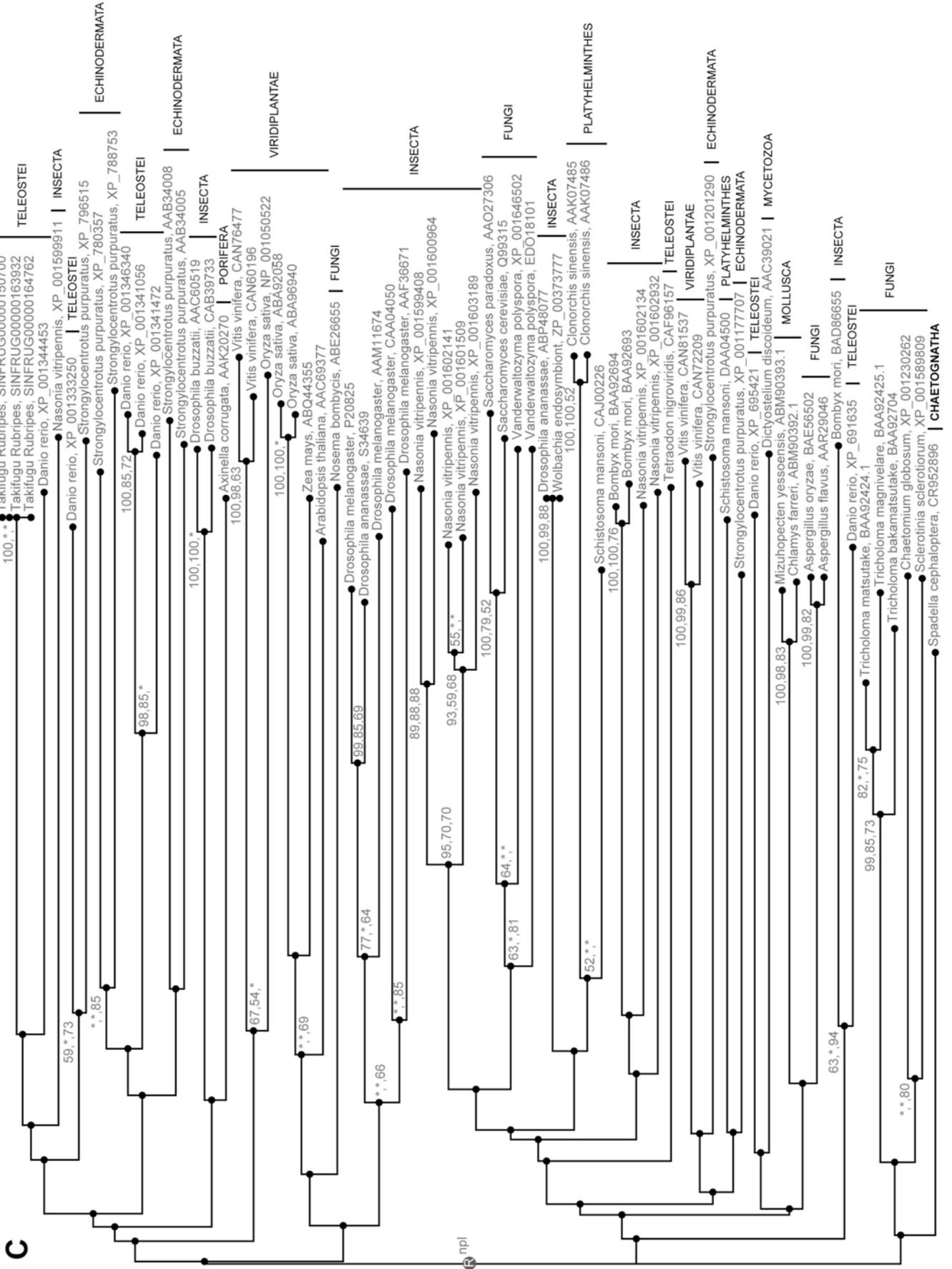
nath RT sequences group with fungi sequences or with fungi and plant sequences, but it is not statistically supported (Fig. 1A and C), whereas in the other analysis, chaetognath sequence is the sister group of a clade containing sequences from fish (Teleostei), echinoderms, plathyhelminthes and insects (Fig. 1B). In this analysis, chaetognath sequences seem to constitute a deep-branching clade of one of the animal retrovirus-like clade. In addition, another phylogenetic analysis suggests that the predicted sequence of two similar ESTs belongs to the *Bel-Pao* superfamily (Fig. 1D) [30]; the chaetognath sequence appears to be the sister group of deuterostomian and plathyhelminth sequences. In all the phylogenetic analyses using LTR-retrotransposons sequences, one or more bacterial sequences are included. They belong to the bacterial endosymbiont *Wolbachia* and exhibit a high homology percentage with the orthologs of their host (*Drosophila ananassae*); the bacterial sequence of the deduced *pol* gene of the *Gypsy* element and of the *Bel-Pao* element share respectively 703 and 1265 strictly contiguous identical amino acid with their insect orthologs. Similarly, in another phylogeny analysis, polymerase sequence of a Pineapple bacilliform virus (Retro-transcribing viruses, Caulimoviridae) [31]) groups with RT sequences from plants (Fig. 1A).

As it has been already shown for several other genes, this supports the hypothesis of horizontal gene transfers [32]. The two last phylogenetic analyses suggest that two chaetognath ESTs could encode part of the RT of a LINE element (Fig. 1E and F). One predicted sequence group LINE elements from a turtle and an annelid, whereas the other groups with nematode and echinoderm sequences; however, in these two analyses the observed groupings are not statistically supported. Moreover, these phylogenetic analyses suggest that the chaetognath genome bears at least two types of LINE elements.

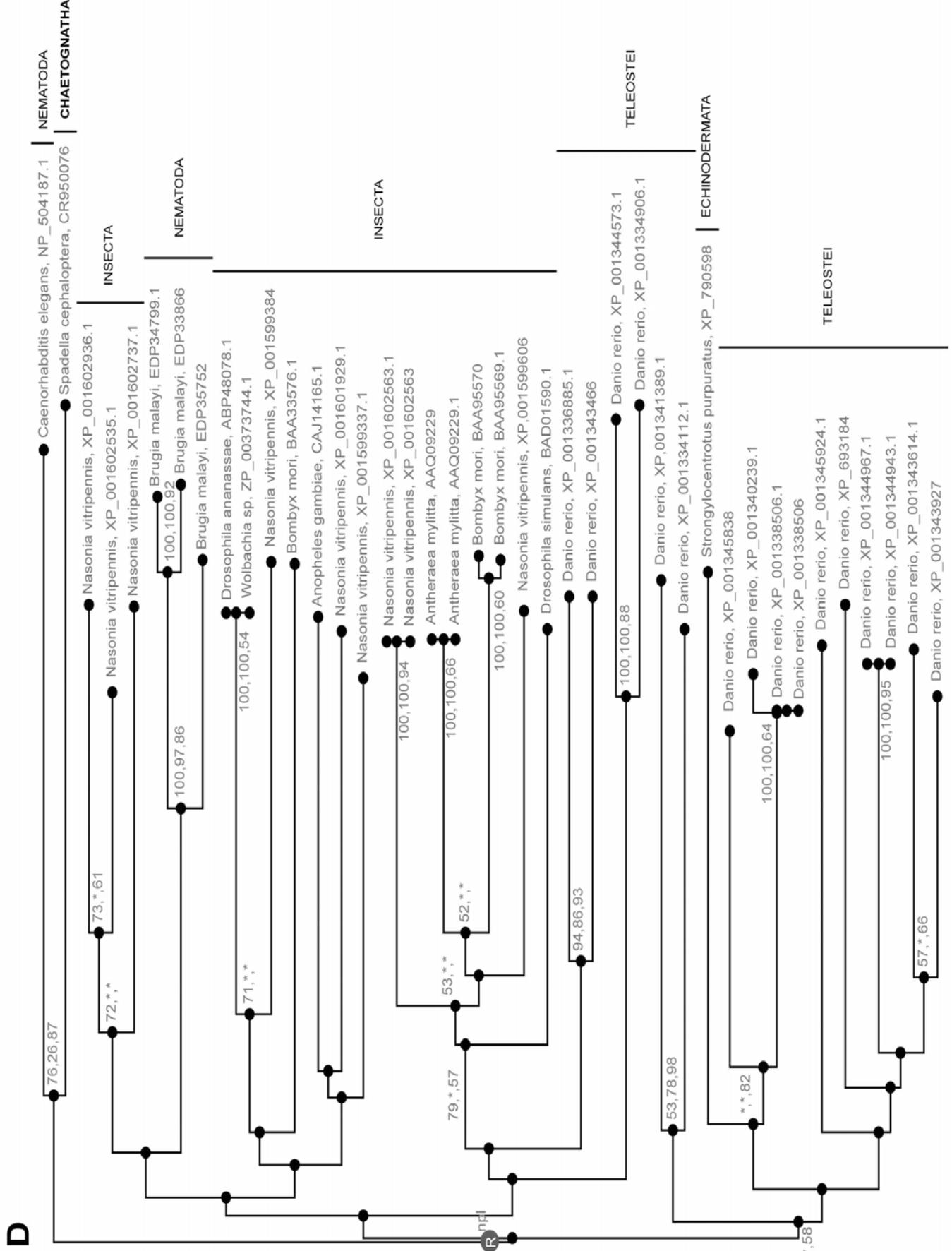
For the other chaetognath EST sequences, due to the low number of homologous sequences, phylogenetic analyses cannot be performed and the predicted amino acid sequences have been aligned with the homologous sequences given using Blast analyses (Fig. 2). The deduced sequences from three ESTs exhibit homologies with parts of the RT domain of some of the *Penelope*-like elements (Fig. 2A, B and C), whereas one EST shares partial homology with the COOH-terminal amino acid region of a putative integrase found in a urochordate (Fig. 2D).



(Fig. 1) contd.....



(Fig. 1) contd.....



(Fig. 1) contd.....

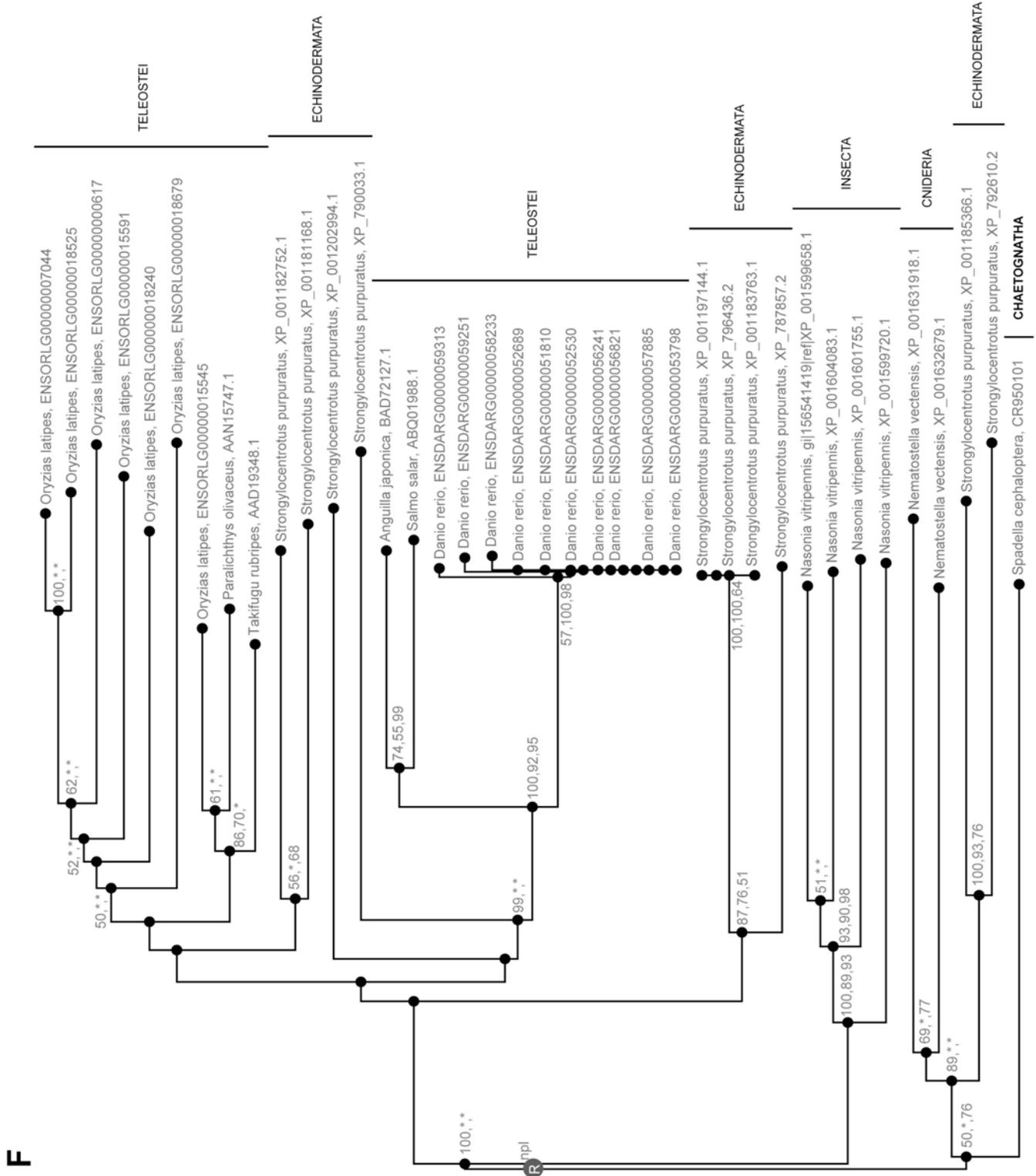


Fig. (1). Phylogenetic trees from predicted translations of coding regions of reverse transcriptase domain of *Gypsy/Ty3* LTR-retrotransposons in (A), (B) and (C), of *Bel-Pao* LTR-retrotransposons in (D) and LINE elements in (E) and (F). The chaetognath EST accession numbers CR953949, (CR953634, CR953418, CR953554), CR952896, (CR950076, CR950075), CR950197 and CR950101 in respectively (A), (B), (C), (D), (E) and (F) phylogenetic analyses. Each tree is the fusion of three phylogenetic trees built on Neighbor Joining (NJ), Maximum Parsimony (MP), and Maximum Likelihood (ML) methods using FIGENIX software platform. For each node, bootstrap values are reported for each method. *, means that the bootstrap value was inferior to 50% (e.g. a bootstrap 69, 57 means that the node exists in NJ tree with a bootstrap value equal to 69, it exists in ML tree with a bootstrap value equal to 57, but does not exists in the MP tree). Branch lengths are correlated to the sequence evolution rate. Many of the same genes are mentioned more than once because they represent different GenBank sequences; since the trees are generated automatically, no discrimination was made.

(A)

CR941783.Chaetog. 1-LSAAGLLAKTTRPATSNTIAKERKALRDMQKDANIILPADKGRATVVMERQEYNEKVES-60
 Teleostei.Tr 228-LKVTAALA-SAKPPTSNTITNEEKRAIASLAKDKNITILPADKGRCTVVLNNTDYDTKILS-286
 Teleostei.Tn 231-LKVSAAALA-NAKPPASNITALEKRALASLAKDKDITILPADKGRCTVVLNNTDYDSKILN-289
 Echinodermata.Sp 94-RRDVDNIIQKARSPKHNI SKEMHTALKTLKADDSITILPADKGRATIIILDTPIYHEKLTID-153
 Platyhelminthes.Sm 131-IRQSIIPLTQRMKGNNQLTSQEQTALKKLRGTGKDIVIVPADKGRRTTVIMDKEEYIKKAE-190
 Insecta.Dv 252-LVKEHKTKNNQNSRDRAILDVEQTRKLLKENINIKILSSDKGNKTVAMDEDEYKKNKMTN-311
 Insecta.Dm 252-LVKEHKTKNNQNSRDRAILDVEQTRKLLKENINIKILSSDKGNKTVAMDEDEYKKNKMTN-311

CR941783.Chaetog. 61-MLSNPVDYTPPLHGDPATAAYQAKIVKCLKPL-90
 Teleostei.Tr 287-LLTDTATYEKLRDPTSSYKKKVVDDLQNL-316
 Teleostei.Tn 290-LLGDSNTYEKLRDPTSTYKKKVIDLLQKL-319
 Echinodermata.Sp 154-LLES-GPYRVLLKDPDRLARKLTNTLLNL-182
 Platyhelminthes.Sm 191-ILEDKSTYKPMIDINPVKKLDNRISKTLNKL-220
 Insecta.Dv 312-ILDDLCAVRTLRDPTSRLOTKNNTFVAQL-341
 Insecta.Dm 312-ILDDLCAVRTLRDPTSRLOTKNNTFVAQL-341

CR941783.Chaetog. 118-WYRIADNTITAIIPGDQVEEFHQHLNSLHLRDKWTKEEEGRLHPXAXWV-----PQY-170
 Teleostei.Tr 508-WFRYVDDTWVKIQIQELEAFSDHLNKTDEHVKFTREEVKGNSLAFLDCAVKITEDRNLT-567
 Teleostei.Tn 511-WFRYVDDTWVKIQTKVEKFTTHLNQTDTFVKFTREDVKGNSLAFLDCEIRIEEDRNLSI-570
 Echinodermata.Sp 376-WKRYVDDTFIIVBISSETDKLLKHMSLEPTIQFTSEIECEGKIAFLDRTLVRHHDNHLST-435
 Platyhelminthes.Sm 409-WIRHVDDTFLIVKKEELEHTYKLINNVFNDIKFTMKQESNDKLPFLDILITRNTKPKPET-468
 Insecta.Dv 531-LTKYVDDLFAITNKIDVENILKELNSFHQKIKFTMELEKDGKLPFLDSIVSRMDNTLKIK-590
 Insecta.Dm 531-LTKYVDDLFAITNKIDVENILKELNSFHQKIKFTMELEKDGKLPFLDSIVSRMDNTLKIK-590

CR941783.Chaetog. 171-SVYRKKTHTDQLQQ-184
 Teleostei.Tr 568-EVYRKPPTHTDQYLQ-581
 Teleostei.Tn 571-EIYRKPPTHTDQYLL-584
 Echinodermata.Sp 436-SVYRKPPTHTDQYIA-449
 Platyhelminthes.Sm 469-QVYRKPPTHTDQILN-482
 Insecta.Dv 591-W-YRKPIASGRILN-603
 Insecta.Dm 591-W-YRKPIASGRILN-603

(B)

CR950255.Chaetog. 1-DIQRKMIFKYISNKSLA---LDVAIRHLRPNGLTSLFSVYWKPTHTDQYIHFERHVPLAHKL-57
 Teleostei.Tr 535-DEHVKFTREEVKGNSLAF--LDCAVKITEDRNLTIEVYRKPPTHTDQYLOFDSHHPLEHKL-592
 Echinodermata.Sp 403-EPTIQFTSEIECEGKIAF--LDTLVHRHDNHLSTSVYRKPPTHTDQYIAFDSHHPISVKR-460
 Platyhelminthes.Sm 436-FNDIKFTMKQESNDKLPF--LDILITRNTKPKPETQVYRKPPTHTDQYILNYSNSPRTHKV-493
 Insecta.Aa 56-SVHKDIKFTHEEEKDGKLPFLDLLVTRREESSTYNFEIYRKPNTQRVIPYTSNHSFQHKM-115

CR950255.Chaetog. 58-ATVRSLTRRAALIPSTQRQRDEELARVRKALSINGYPKWAFEAGKHR---PRHQAEPTKT-114
 Teleostei.Tr 593-GVIRTLQHRAREIPTTSQGRKKEQNHIKTAALKTCGYPDWAFK-----T-636
 Echinodermata.Sp 461-GLVKCLFHRASRIVTSPQPCKEBTRVRSALSINGYPRRFIHNTKNRSSEPRYQKVYKTF-520
 Platyhelminthes.Sm 494-NCVHTLFKRRARTHCTIIPAQLYVVLYLKTTLQKNGYLINFIKKYQPH---PSSEIKSTTE-550
 Insecta.Aa 116-AAFHMIHRMQTLPLSEDGKTKLELEYIYETARINGYKERTIKAIIDK---KERQORIRNAL-172

CR950255.Chaetog. 115-TNDDINK-122
 Teleostei.Tr 637-SRKQDLSK-644
 Echinodermata.Sp 521-TVLPYIDG-528
 Platyhelminthes.Sm 551-INKRITIP-558
 Insecta.Aa 173-TLTPITE-180

(Fig. 2) contd.....

(C)

```

CR950254.Chaetog. 1-QQRQYERYPISVP-----YHDSISQPI TRAMHKAGLXTYVKPRGL-LREALVHVPKDK-51
Teleostei.Tr      646-EEERNKRRSVSIIP-----YLSGVSEKFRRILOKHDIPVHFKPSNT-LRQRLVHVPKDK-696
Echinodermata.Sp 512-RYQKVYKTFVLP-----YIDGVSQQLKRRLESHGIRTVFRSDTT-LRQQLTRPKDP-562
Platyhelminthes.Sm 546-KSTTEINKRITIP-----YIKGISETKRLLKAFRINVAHKPTKS-LHSILCKPKDE-596
Insecta.Aa       176-TPITEPTKRVSIIP-----YDVHISKQLRPKLRNFGIDLVSFSSRDNLRTSLGSTKDP-227
Insecta.Dv       659-KRRQIERKKPTEPAKIYKSLIYVPRLSERLTNSDCYNKQDIKVAHKPT-NTLQKFFNKIK-717
Insecta.Dm       659-KRRQIERKKPTEPAKIYKSLIYVPRLSERLTNSDCYNKQDIKVAHKPT-NTLQKFFNKIK-717

CR950254.Chaetog. 52-VDTERETRVIIYWMASAGV---NSAPCNAAEHGETERALRDRTAEHLAATRA PLLFNSAI-108
Teleostei.Tr      697-TPRPKQSNVYAVQCQ-----EKCKELYIGETKQPLHRRMAQHRRATSSGQ---DSAV-746
Echinodermata.Sp 563-IPPHRRDGI VYNI PCK-----GCDRSYIGETARPIKERITEHKRDVRLRR-TDNSAV-613
Platyhelminthes.Sm 597-ITKEDKPNIIYKINCA-----NCNKHYIGQSGRPLHLRLQEHQLAVKHH--DMSSLI-646
Insecta.Aa       228-VNT-----LNKAGVYKISCSHC SKVYVGOTKRSLEVRFEHLAEIGKAOKTIDKGM-278
Insecta.Dv       718-SKIPMIEKSNVVYQIPCGGDNN-NKCNSVYIGTTKSKLKTRISQHKSDFKLRHQNNIQKT-776
Insecta.Dm       718-SKIPMIEKSNVVYQIPCGGDNN-NKCNSVYIGTTKSKLKTRISQHKSDFKLRHQNNIQKT-776

CR950254.Chaetog. 109-IQHATDNHRHFRCGDARVLCRD-130
Teleostei.Tr      747-HLHLKESGHSFEDSQVRILARE-768
Echinodermata.Sp 614-AEHAWDNQHQPDWDGVCLSQE-635
Platyhelminthes.Sm 645-SMHVDNYRHTFDWKNVEILDREG-666
Insecta.Aa       279-TYDFKSKVAEHIFSEGHTITTA-300
Insecta.Dv       777-ALMTHCIRSNTNPNFDETTILQ-798
Insecta.Dm       777-ALMTHCIRSNTNPNFDETTILQ-798

```

(D)

```

CR950196.Chaetog. 2-GQSMGKTYLAAKLIKYKEEMFSPVPKRVIIYTYMEYND---ISKDLANGYGVFEYEGY-K-57
Urochordata.Od   25-GSSGAGKTFIAKQILIRH-LFQKKT KRVVYFYPCYLDEKPVNWDLEDLGIPI TYRVGIPS-83

CR950196.Chaetog. 58-QAVVDSINPSYCTMLVLDDFLNEKNVHKQIVDLFIKSVHKKNLCLIFMVQNLFYKGF PNL-117
Urochordata.Od   84-QDDIDVMLPH--TIVLDDTYDEAIASTAI DHLFRVNSGKKLLNVLIMTQNNFAAG-----137

CR950196.Chaetog. 118-RSISLDSQRMFLGRSIRDRQNI VTLARQSMPNRV-----PFFLESYELATK----EPY-166
Urochordata.Od   138-----KYGRNIRNNCNITILLRNCLDTRINKSVCSQAGLLKAYQRAKEDLKNQSY-187

CR950196.Chaetog. 167-SFLLFD CDQRQKNEELRLRSNIFPDET TV-195
Urochordata.Od   188-PYMFLDQSPRAHASGYRLYTNIFE EYPTV-216

```

Fig. (2). Alignments of predicted amino acid sequence of chaetognath ESTs with the homologous regions of various retrotransposons. (A), (B), (C) and (D) correspond respectively to alignments with CR941783, CR950255, CR950254, CR950196 chaetognath ESTs. In each alignment, the first sequence corresponds to the predicted amino acid sequence of a chaetognath EST. In (A), (B) and (C) the reverse transcriptase homologous sequences belong to Penelope-like elements, whereas in (D), only one putative integrase has been found partially homologous to chaetognath sequence. Sequences were aligned using program CLUSTAL W; positions of nucleotides are shown. The abbreviations of the other sequences and their accession numbers are: (Teleostei.Tr) *Takifugu rubripes*, AAK58879.1; (Teleostei.tn) *Tetraodon nigroviridis*, CAF94542; (Echinodermata.Sp) *Strongylocentrotus purpuratus*, XP_001200565; (Platyhelminthes.Sm) *Schistosoma mansoni*, DAA00890; (Insecta.Aa) *Aedes aegypti*, AAZ15235; (Insecta.Dv) *Drosophila virilis* AAL14979; (Insecta.Dm) *Drosophila melanogaster*, AAR86938. Amino acid residues which are identical to those of the chaetognath sequences are underlined in yellow. Dashes represent gaps introduced to improve the sequence alignment. Stop codons appear as a result of unknown amino acid residue marked as X.

Closed Relationships Between Retrotransposons and Microsatellites

Analysis of chaetognath ESTs reveals that two sequences contain part of retrotransposable elements stringly associated with the same microsatellite repeat motif (CAA)_n (Table 2). This motif has already been found in two out of the four microsatellite loci known of another chaetognath, *Sagitta*

setosa (DQ463218 and DQ463220) [33]. However, in this last species, no MGEs have been found in the flanking sequences; the first loci is at the end of the ribosomal protein L8 gene, and the second in an intron of the Midasin gene (or pseudogene). Interestingly, several microsatellite families exhibit MGEs in their flanking sequences both in plants and animals [34-36] suggesting a possible close relationships between these two types of repetitive sequences.

Table 2. Analysis of the Chaetognath ESTs Containing Telomeric Motifs

Number of Repeat	1	2	3	4	5	6	7
Motifs							
TTAGGG "Metazoan"	36	10	-	2	-	-	1
TTAGG "Arthropoda"	24	1	1	-	-	-	-
TTAGG/TTAGGG or TTAGGG/TTAGG	3	-	-	-	-	-	-
TTAGGG/TTAGGG/TTAGG	1	-	-	-	-	-	-
TTAGG/TTAGG/TTAGGG	1	-	-	-	-	-	-

The accession numbers of these ESTs are: CR940792, CR941840, CR944786, CR945111, CR945171, CR946582, CR946615, CR946633, CR946651, CR946666, CR947615, CR948211, CR949136, CR949498, CR949663, CR949907 and CR950054. - as 0.

Research of Retrotransposons and Inverse Transcriptase Genes in Telomeric Regions

Telomeres are regions of repetitive DNA at the end of most of the linear chromosomes, which protects the end of the chromosome from destruction [37]. Telomerase, a ribonucleoprotein composed of a reverse transcriptase (RT) and an RNA encoded by a different gene, synthesizes the telomeric DNA repeats [38]. A relationship between telomerases and retrotransposon RTs, which are encoded by their template RNA, was surmised from shared amino acid motifs [38,39] and the observation that telomeres are elongated by non-LTR retrotransposons in some insects [40]. Indeed, as in almost all studied organisms, telomere repeats are very short simple sequences the exceptions are dipterans, which evolved chromosome ends with complex arrays of long satellite repeats in both *Chironomus* and *Anopheles gambiae* [41,42] or of non-LTR retrotransposons (members of *Drosophila* genus) that are elongated by telomerase-independent mechanisms (for review see [43,44]). *Drosophila melanogaster* telomeres are composed of two non-LTR retrotransposons, *HeT-A* and *TART*, with few copies of *Tahre*, an element that appears to combine parts of both *HeT-A* and *TART* and homologs of these two last elements have been found in *D. virilis* [45]. Moreover, one of the chaetognath EST sequences exhibits homology with the RT domain of the *Penelope*-like elements (Table 1 and Fig. 1D). This order of retrotransposon has been found in animals, fungi and plants; in some taxa, they can be associated with telomeric repeats as shown by Gladyshev and Arkhipova [46]. According to these authors, they may descend from the missing link between early eukaryotic retroelements and present-day telomerases.

As a large number of ESTs have been found in telomeric regions of both in animals and plants (for example [47,48], consensus motifs of various taxa including animals, protozoa, plants and fungi [49] have been found in the chaetognath EST collection. ESTs containing telomeric-like sequences have been researched using all the known telomeric sequences. Only ESTs containing at least three telomeric motifs in a sequence of 153 nucleotides have been selected. Using this non stringent criterion, seventeen ESTs have been found and all but two contain both TTAGGG and TTAGG motifs (Table 2). So (TTAGGG)_n could be the ancestral te-

lomere repeat motif of Metazoa. It has been conserved from the metazoan radiation in most animal phylogenetic lineages and according to our knowledge replaced by other motifs, only in two major lineages, Arthropoda [(TTAGG)_n] and Nematoda [(TTAGGC)_n] [50]. In the chaetognath EST collection, the mean nucleotide distance between two motifs is 15 nt and 19 motifs in tandem or more have been found. The nematode TTAGGC motif has very rarely been found but never in tandem or more, suggesting probably it is a TTAGG motif with a C as flanking base in 3' end or a variant of the metazoan consensus.

Moreover, in addition to their role in protecting the ends of chromosomes, telomeres also influence the expression of adjacent genes, a process called telomere-position effect. The pattern of expression of the telomeric transgenes demonstrates that subtelomeric regions are epigenetically reprogrammed [51]. However, when the seventeen sequences were translated in all six reading frames and compared to the sequences in protein databases, the result has been negative indicating that these regions are not closely located to genes. A recent study suggests that in *Drosophila* an existing non-LTR retrotransposon was recruited to perform the cellular function of telomere maintenance [52]. However, in our analysis, the telomere flanking regions do not contain sequence homology with any retrotransposon element, and never even *Penelope*-like element, suggesting the presence of only consensus telomere repeats in chaetognath.

DISCUSSION

Present analysis of a chaetognath EST library allowed to find thirteen retrotransposon sequences. Retrotransposons comprise 0.11% and 0.54 % of respectively the total ESTs and of the number of ESTs assigned to protein gene transcripts. In other animals, the corresponding values are generally higher, e.g., 1.0% in venom glands of a fish [53], from 4 to 14 according to the stages in a platyhelminth [54], and 14% in mouse oocytes [55]. However, the chaetognath value is similar to the 0.12% reported from the survey of ESTs from many plant species [56] but higher values could be found [57]. These transcriptional differences could be due to severe controls of the retrotransposon expression in higher plants, in which LTR-retrotransposons alone can comprise 50–90% of the genome [58] and probably also in juvenile chaetognaths. However, EST collection can exhibit several bias; indeed, ESTs are short single-pass sequence reads derived from cDNA clones selected randomly from cDNA libraries, and in contrast to genomic sequences, they are generally of low quality, poorly annotated, and highly redundant. Common EST features include ~2% sequence error rate with high frequency of insertions and deletions, redundancy of sequences derived from highly expressed genes, and low representation of genes expressed at low levels. Moreover, ESTs may derive from unspliced immature mRNAs, alternative splicing and polyadenylation sites, cloning artifacts (chimerisms), and mitochondrial transcripts [59]. In addition, it should be kept in mind that only transcription and ultimate integration in tissues giving rise to gametes is heritable and the chaetognath EST collection was derived from whole animals.

In spite the low number of ESTs encoding retrotransposons found in the chaetognath library studied, members of

three out of four orders of autonomous retrotransposons have been found [21]; the lacking order is the one containing the *DIRS*-like elements. The higher number of transcripts has been found for the *Gypsy* retrovirus-like superfamily which could constitute a deep-branching clade according to phylogenetic analyses. The origin of chaetognaths remains obscure, but fossil evidences suggest that this phylum was widespread and diverse in the earliest Cambrian [60], and the difficulties of the phylogenetic position of this taxon is probably partly due to its divergence at an early stage from the primitive ancestor of the Bilateria. Moreover, as it is well known [61], present retrovirus-like phylogenies suggest the polyphyly of these elements.

In addition, one deduced amino acid sequence of two similar ESTs exhibits sequence homology RT of *Bel-Pao* LTR-retrotransposons issuing from deuterostomian and platyhelminthes. Moreover, phylogenies suggest the presence of active LINES in the chaetognath genome; however, the level of homology between these sequences and the closer published sequences is very low. Phylogenetic analyses of LINE-like RT sequences suggested that sequences of the chaetognath *Sagitta* sp. group with those of Lophotrochozoa (i.e., Nemertea, Mollusca, Gastrotricha, Annelida, Echiura and Rotifera) [16], whereas in our phylogenies, the chaetognath LINE elements group with vertebrae and insect sequences.

Alignments of other predicted EST sequences with the closer orthologs obtained using Blast analyses show that the chaetognath genome also contains active elements which exhibit sequence homology with RT of *Penelope*-like elements issuing from deuterostomian, platyhelminthes and insects [39,62,63]. However, all our analyses show that the chaetognath retrotransposon sequences are not phylogenetically informative; this is principally due to the short size of the regions analyzed. Moreover, both in plants and animals numerous evidences of horizontal transmission have been published [64-66]; however, the basal positions of chaetognath sequences in most of our phylogenetic trees seem to exclude horizontal transfers.

Association between microsatellites and MGEs has been reported in a variety of organisms including plants and animals [34-36]. These MGEs could be DNA transposons, but many of them are retroelements. This results in microsatellites that have similar or nearly identical flanking regions that share great homology with MGEs, suggesting these last elements can be involved in the genesis and genomic spread of microsatellites in organisms as diverse as animals [34,35,67-69] and plants [70]. This intimate association between microsatellite repeats and retrotransposons has also been put to good use for develop a method named REMAP for genotyping and fingerprinting analyses [71]. The microsatellite-MGEs association could be a molecular symbiosis between two types of genomic sequences. Indeed, the two main molecular pathways of this mutual aid are by means of transposition and recombination. MGEs can not generally invade some chromosomal regions, due, e.g., to the absence of insertion sites or the presence of euchromatin. Moreover, MGEs can be strongly negatively regulated by various mechanisms, they even can encode their own negative *trans*-regulator [72, 73]. Recombination-related events due to microsatellites, such as unequal crossing over, allow the expan-

sion of MGEs present in the microsatellite flanking regions, even when they are inactive. Contrarily, microsatellites flanking MGEs sequences could be multiplied and dispersed in the genome during transposition processes. Transposition of additional regions including functional genes are known for both DNA transposons for which complex transposons probably evolve by transposition of homologous insertion sequences to nearby sites within a DNA molecule [74], and retrotransposons which can mediate sequence transduction [75]. Moreover, the 3' end region of several non LTR-retrotransposons can be implicated as a major source for formation of adenine-rich microsatellites [67]. The potential molecular symbiosis between MGEs and microsatellites show that the behavior and the evolution of repetitive sequences can only be understood within a larger genomic context.

Our research of ESTs containing telomeric-like sequences reveals the presence of two types of telomeric motifs. The vertebrate motif (TTAGGG) is dominant; it constitutes an ancestral motif of telomeres in bilaterian animals and possibly also in the superclade including animals, fungi and amoebozoans [50,76]. More surprisingly, the probably ancestral Arthropoda motif (TTAGG)_n [76] has also been found in ~30% of the cases, suggesting that it is not a type of degenerate TTAGGG repeats. Moreover, no sequence of retrotransposons or RT genes has been found in the telomeric regions.

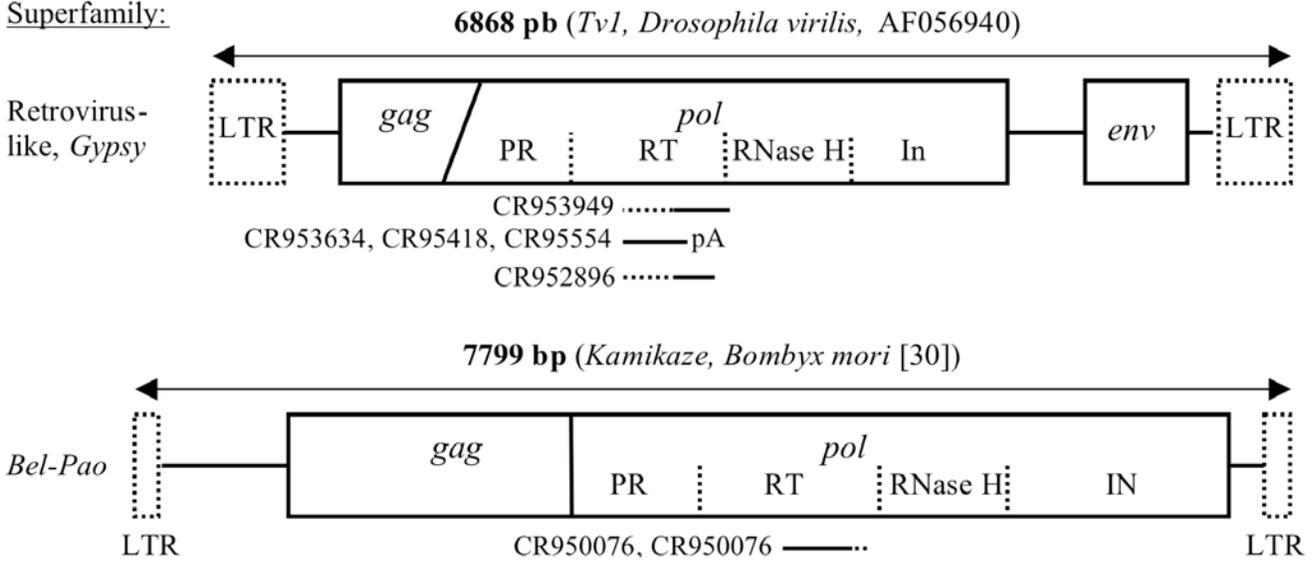
The ability of various factors to stimulate MGE activity was first proposed by McClintock [77], and one of us (EF) has been involved in this field of research for a long time. With regard to LTR retrotransposons and retrovirus, chemical and physical agents have been shown to induce transcription and transposition [78-92]. Non-LTR retrotransposons also respond to these stress [93]. Moreover, expression of retrotransposon promoters are wound-inducible [94,95] and retrotransposons can be activated during viral, bacterial, fungal or parasitic attacks [96-98]. As retrotransposition burst can be an indicator of the stress genomic response, activation of MGEs will be investigated in chaetognaths, because this taxon seems very resistant. Indeed, chaetognaths do acquire eukaryotic and bacterial parasites but not very frequently [99,100]. There is no host-specific parasite known in chaetognaths, which is remarkable for such an old group [60]. Moreover, chaetognaths exhibit a great antibacterial activity [101] and in laboratory, no beheaded chaetognaths, which can survive some 30 days after decapitation and are able to mature spermatozoans and to mate with normal specimens; have exhibited bacterial infestation [102].

CONCLUSION

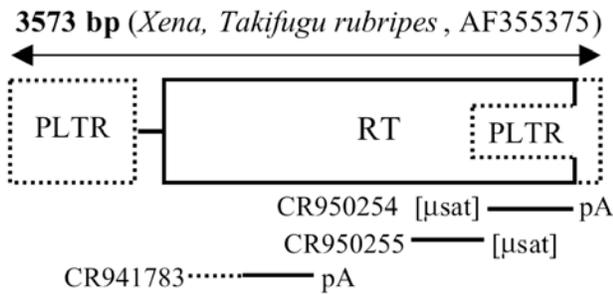
In spite the low number of retrotransposon ESTs found in a juvenile chaetognath collection added to the small size of the sequences analyzed, this study suggests that chaetognath retrovirus-like retrotransposons could constitute deep-branching clades. The origin of these elements could correspond to the origin of the phylum; indeed, fossils of chaetognath grasping spines support the hypothesis that this animal taxon was present in the Cambrian times or even earlier. Moreover, studies on chaetognath retrotransposons could highlight future research in the exciting domain of the evolutionary origin of the retrovirus. Lastly, owing to the role of

LTR-retrotransposon

Superfamily:



Penelope-like element (Ple)



LINE

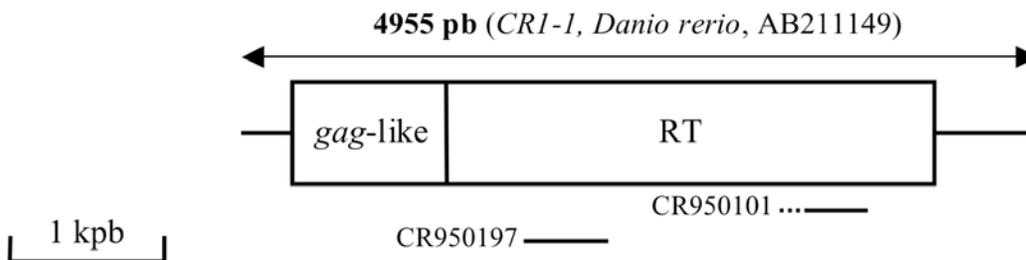


Fig. (3). Mapping of the chaetognath ESTs bearing retrotransposon sequences on one of the closer complete retrotransposon genome. For each EST, homologous region between the retrotransposon is in continuous line, whereas, region which does not exhibit homology is in dash line. Concerning the retrotransposons, open reading frames are surrounded by continuous lines, non coding region are symbolized by a continuous line, whereas, long terminal repeats (LTR) or “pseudo-LTR” structure (PLTR) are surrounded by dash lines. For each retrotransposon, the length in base pairs (bp), its name, the name of the species bearing the element and the accession number or the corresponding reference are given. Abbreviations: [usat] region containing microsatellite repeats, (pA) polyA trail, (*gag*) group antigen gene, (*pol*) polymerase gene, (*env*) envelope gene; and the various domains of the *pol* gene: (PR) protease, (RT) transcriptase inverse, (RNase H) RNase H, (IN) integrase.

LTR-retrotransposons on genome structure, evolution and function, entire elements will be cloned and characterized, starting from the retrotransposon fragments shown in this

study. Moreover, activation of these elements during normal development and in situations of stress, including pathogen attack will be sought.

ABBREVIATIONS

MGE	=	Mobile Genetic Element
DNA	=	Deoxyribonucleic Acid
RNA	=	Ribonucleic Acid
RT	=	Reverse Transcriptase
LTR	=	Long Terminal Repeat
Gag	=	Group antigen
Pol	=	Polymerase
PR	=	Proteinase
IN	=	integrase
PLE	=	<i>Penelope</i> -like element
ORF	=	Open Reading Frame
RNase H	=	Ribonuclease H
ERV	=	Endogenous Retrovirus
EST	=	Expressed Sequence Tag
BP	=	Bootstrap value
ENV	=	Envelope

REFERENCES

- Feigenbaum DL, Maris RC. Feeding in chaetognatha. *Oceanogr Mar Biol Ann Rev* 1984; 22: 343-392.
- Casanova J-P. Chaetognatha. In: Boltovskoy D, Ed. *South Atlantic Zooplankton*. Leiden: Backhuys Publishers 1999; 1353-1374.
- Lewbart AL. *Invertebrate Medicine*. New York: Wiley, John & Sons; 2004.
- Shimotori T, Goto T. Developmental fates of the first four blastomeres of the chaetognath *Paraspadella gotoi*: relationship to protostomes. *Dev Growth Differ* 2001; 43(4): 371-382.
- Faure E, Casanova J-P. Comparison of chaetognath mitochondrial genomes and phylogenetical implications. *Mitochondrion* 2006; 6(5): 258-262.
- Marletaz F, Martin E, Perez Y, *et al.* Chaetognath phylogenomics: a protostome with deuterostome-like development. *Curr Biol* 2006; 16(15): R577-578.
- Matus DQ, Copley RR, Dunn CW, *et al.* Broad taxon and gene sampling indicate that chaetognaths are protostomes. *Curr Biol* 2006; 16(15): R575-576.
- Casanova J-P, Duvert M, Perez Y. Phylogenetic interest of the chaetognath model. *Mésogée* 2001; 59: 27-31.
- Telford MJ, Holland PWH. Evolution of 28S ribosomal DNA in Chaetognaths: duplicate genes and molecular phylogeny. *J Mol Evol* 1997; 44(2): 135-144.
- Papillon D, Perez Y, Caubit X, Le Parco Y. Systematics of Chaetognatha under the light of molecular data, using duplicated ribosomal 18S DNA sequences. *Mol Phyl Evol* 2006; 38(3): 621-634.
- Barthélémy R-M, Chenuil A, Brancart S, Casanova J-P, Faure E. Translational machinery of the chaetognath *Spadella cephaloptera*: A transcriptomic approach to the analysis of cytosolic ribosomal protein genes and their expression. *BMC Evol Biol* [serial on the Internet]. Aug 2007; [cited 2007 August 28] 7:146 [about 45 screens]. Available from: <http://www.biomedcentral.com/1471-2148/7/146>
- Barthélémy R-M, Péténian F, Vannier J, Casanova J-P, Faure E. Evolutionary history of the chaetognaths inferred from actin and 18S-28S rRNA paralogous genes. *Int J Zool Res* 2006; 2(4): 284-300.
- Barthélémy R-M, Grino M, Pontarotti P, Casanova J-P, Faure E. The differential expression of ribosomal 18S RNA paralog genes from the chaetognath *Spadella cephaloptera*. *Cell Mol Biol Lett* 2007; 12(4): 573-583.
- Barthélémy R-M, Casanova J-P, Grino M, Faure E. Selective expression of two types of 28S rRNA paralogous genes in the chaetognath *Spadella cephaloptera*. *Cell Mol Biol (Noisy-le-grand)* 2007; 53(Suppl:OL): 989-993.
- Barthélémy R-M, Grino M, Pontarotti P, Casanova JP, Faure E. A possible relationship between the phylogenetic branch lengths and the chaetognath rRNA paralog gene functionalities: ubiquitous, tissue-specific or pseudogenes. In: Pontarotti P, Ed. *Evolutionary biology from concept to application*. New-York, Spinger; 2008; in press.
- Arkhipova I, Meselson M. Transposable elements in sexual and ancient asexual taxa. *Proc Natl Acad Sci USA* 2000; 97(26): 14473-14477.
- Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA* 1997; 94(15): 7704-7711.
- Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science* 2004; 303(5664): 1626-1632.
- Labrador M, Corces VG. Interactions between transposable elements and the host genome. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, Eds. *Mobile DNA II*. Washington, DC, American Society for Microbiology Press 2002; 1008-1023.
- Levy A, Sela N, Ast G. TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. *Nucleic Acids Res* 2008; 36: D47-D52.
- Wicker T, Sabot F, Hua-Van A, *et al.* A unified classification system for eukaryotic transposable elements. *Nature Rev Genet* 2007; 8(12): 973-982.
- Bucheton A. The relationship between the *flamenco* gene and *gypsy* in *Drosophila*: how to tame a retrovirus. *Trends Genet* 1995; 11(9): 349-353.
- Arkhipova IR. Distribution and phylogeny of Penelope-like elements in eukaryotes. *Syst Biol* 2006; 55(6): 875-885.
- Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 2003; 35(1): 41-48.
- Brosius J. Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 1999; 107(1-3): 209-238.
- Ding W, Lin L, Chen B, Dai J. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB Life* 2006; 58(12): 677-685.
- Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EGJ. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. *BMC Bioinformatics* [serial on the Internet]. Aug 2005; [cited 2005 August 5] 6: 198 [about 25 screens]. Available from: <http://www.biomedcentral.com/1471-2105/6/198>
- Larkin MA, Blackshields G, Brown NP, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; 23(21): 2947-2948.
- Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 1990; 9(10): 3353-3362.
- Abe H, Ohbayashi F, Sugasaki T, *et al.* Two novel Pao-like retrotransposons (Kamikaze and Yamato) from the silkworm species *Bombyx mori* and *B. mandarina*: common structural features of Pao-like elements. *Mol Genet Genomics* 2001; 265(2): 375-385.
- Thomson KG, Dietzgen RG, Thomas JE, Teakle DS. Detection of pineapple bacilliform virus using the polymerase chain reaction. *Ann Appl Biol* 1996; 129(1): 57-69.
- Salzberg SL, Hotopp JC, Delcher AL, *et al.* Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* [serial on the Internet]. Feb 2005; [cited 2005 February 22] 6(3): R23 [about 18 screens]. Available from: <http://genome-biology.com/2005/6/3/R23>
- Peijnenburg KT, Fauvelot C, Breeuwer JA, Menken SB. Spatial and temporal genetic structure of the planktonic *Sagittia setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Mol Ecol* 2006; 15(11): 3319-3338.
- Megléc E, Peténian F, Danchin E, D'Acier AC, Rasplus JY, Faure E. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Mol Ecol* 2004; 13(6): 1693-1700.
- Megléc E, Anderson SJ, Bourguet D, *et al.* Microsatellite flanking region similarities among different loci within insect species. *Insect Mol Biol* 2007; 16(2): 175-185.

- [36] Zhang DX. Lepidopteran microsatellite DNA: redundant but promising. *TREE* 2004; 19(10): 507-509.
- [37] Nakamura TM, Cech TR. Reversing time: origin of telomerase. *Cell* 1998; 92(5): 587-590.
- [38] Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 1997; 276(5312): 561-567.
- [39] Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB. Retroelements containing introns in diverse invertebrate taxa. *Nat Genet* 2003; 33(2): 123-124.
- [40] Pardue ML, Danilevskaya ON, Traverse KL, Lowenhaupt K. Evolutionary links between telomeres and transposable elements. *Genetica* 1997; 100(1-3): 73-84.
- [41] Rosén M, Edström J. DNA structures common for chironomid telomeres terminating with complex repeats. *Insect Mol Biol* 2000; 9(3): 341-347.
- [42] Walter MF, Bozorgnia L, Maheshwari A, Biessmann H. The rate of terminal nucleotide loss from a telomere of the mosquito *Anopheles gambiae*. *Insect Mol Biol* 2001; 10(1): 105-110.
- [43] Biessmann H, Mason JM. Telomerase-independent mechanisms of telomere maintenance. *Cell Mol Life Sci* 2003; 60(11): 2325-2333.
- [44] Pardue ML, DeBaryshe PG. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu Rev Genet* 2003; 37: 485-511.
- [45] Casacuberta E, Marín FA, Pardue ML. Intracellular targeting of telomeric retrotransposon Gag proteins of distantly related *Drosophila* species. *Proc Natl Acad Sci USA* 2007; 104(20): 8391-8396.
- [46] Gladyshev EA, Arkhipova IR. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc Natl Acad Sci USA* 2007; 104(22): 9352-9357.
- [47] Ning Y, Xu JF, Li Y, *et al.* Telomere length and the expression of natural telomeric genes in human fibroblasts. *Hum Mol Genet* 2003; 12(11): 1329-1336.
- [48] Anderson LK, Lai A, Stack SM, Rizzon C, Gaut BS. Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. *Genome Res* 2006; 16(1): 115-122.
- [49] Podlevsky JD, Bley CJ, Omana RV, Qi X, Chen J. The Telomerase Database. *Nucleic Acids Res* 2008; 36: D339-D343.
- [50] Traut W, Szczepanowski M, Vítková M, Opitz C, Marec F, Zrzavý J. The telomere repeat motif of basal Metazoa. *Chromosome Res* 2007; 15(3): 371-382.
- [51] Gao Q, Reynolds GE, Innes L, *et al.* Telomeric transgenes are silenced in adult mouse tissues and embryo fibroblasts but are expressed in embryonic stem cells. *Stem Cells* 2007; 25(12): 3085-3092.
- [52] Villasante A, Abad JP, Planelló R, Méndez-Lago M, Celniker SE, de Pablos B. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res* 2007; 17(12): 1909-1918.
- [53] Magalhães GS, Junqueira-de-Azevedo JL, Lopes-Ferreira M, Lorenzini DM, Ho PL, Moura-da-Silva AM. Transcriptome analysis of expressed sequence tags from the venom glands of the fish *Thalassophryne nattereri*. *Biochimie* 2006; 88(6): 693-699.
- [54] DeMarco R, Kowaltowski AT, Machado AA, *et al.* Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. *J Virol* 2004; 78(6): 2967-2978.
- [55] Evsikov AV, Graber JH, Brockman JM, *et al.* Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes Dev* 2006; 20(19): 2713-2727.
- [56] Vicent CM, Jääskeläinen MJ, Kalendar R, Schulman AH. Active retrotransposons are a common feature of grass genomes. *Plant Physiol* 2001; 125(3): 1283-1292.
- [57] Rotter D, Bharti AK, Li HM, *et al.* Analysis of EST sequences suggests recent origin of allotetraploid colonial and creeping bentgrasses. *Mol Genet Genomics* 2007; 278(2): 197-209.
- [58] Du C, Swigonová Z, Messing J. Retrotranspositions in orthologous regions of closely related grass species. *BMC Evol Biol* [serial on the Internet]. Aug 2006; [cited 2006 August 16] 6:62 [about 15 screens]. Available from: <http://www.biomedcentral.com/1471-2148/6/62>.
- [59] Gissi C, Pesole G. Transcript mapping and genome annotation of ascidian mtDNA using EST data. *Genome Res* 2003; 13(9): 2203-2212.
- [60] Vannier J, Steiner M, Renvoisé E, Xu S-X, Casanova J-P. Predator arrow worms in the early cambrian food webs. *Proc R Soc B* 2007; 274(1610): 627-633.
- [61] Volff JN, Körting C, Altschmied J, *et al.* Jule from the fish *Xiphophorus* is the first complete vertebrate *Ty3/Gypsy* retrotransposon from the *Mag* family. *Mol Biol Evol* 2001; 18(2): 101-111.
- [62] Dalle Nogare DE, Clark MS, Elgar G, Frame IG, Poulter RT. *Xena*, a full-length basal retroelement from tetraodontid fish. *Mol Biol Evol* 2002; 19(3): 247-255.
- [63] Pyatkov KI, Shostak NG, Zelentsova ES, *et al.* Penelope retroelements from *Drosophila virilis* are active after transformation of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 2002; 99(25): 16150-16155.
- [64] Jordan IK, Matyunina LV, McDonald JF. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc Natl Acad Sci USA* 1999; 96(22): 12621-12625.
- [65] Terzian C, Ferraz C, Demaille J, Bucheton A. Evolution of the *Gypsy* endogenous retrovirus in the *Drosophila melanogaster* subgroup. *Mol Biol Evol* 2000; 17(6): 908-914.
- [66] Eickbush TH, Malik HS. Origins and evolution of retrotransposons. In: Craig *et al.* Eds. *Mobile DNA II*. Washington DC, ASM Press 1999; 1111-1144.
- [67] Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA. *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* 1995; 29(1): 136-144.
- [68] Nadir E, Margalit H, Gallily T, Ben-Sasson SA. Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 1996; 93(13): 6470-6475.
- [69] Wilder J, Hollocher H. Mobile elements and the genesis of microsatellites in dipterans. *Mol Biol Evol* 2001; 18(3): 384-392.
- [70] Ramsay L, Macaulay M, Cardle L, *et al.* Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J* 1999; 17(4): 415-425.
- [71] Kalendar R, Schulman AH. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat Protoc* 2006; 1(5): 2478-2484.
- [72] Faure E. A sequence of the U5 region of *Drosophila* 1731 retrotransposon long terminal repeat (LTR) trans-represses the LTR-directed transcription. *Biochemistry (Mosc)* 1999; 64(6): 678-692.
- [73] Faure E. Strategy of coevolution between an intracellular parasite (retrotransposon) and his host (fruit fly). A sequence of the U5 region of *drosophila* 1731 retrotransposon long terminal repeat (LTR) trans-represses the LTR-directed transcription. *Mésogée* 2001; 59: 7-15.
- [74] Kidwell MG. Transposable elements. In: Gregory TR, Ed. *The Evolution of the Genome*. San Diego, Elsevier 2005; 165-221.
- [75] Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci USA* 2006; 103(47): 17608-17613.
- [76] Vítková M, Král J, Traut W, Zrzavý J, Marec F. The evolutionary origin of insect telomeric repeats, (TTAGG)_n. *Chromosome Res* 2005; 13(2): 145-156.
- [77] McClintock B. The significance of responses of the genome to challenge. *Science* 1984; 226(4676): 792-801.
- [78] Faure E, Best-Belpomme M, Champion S. Upregulation of the *Drosophila* 1731 retrotransposon long-terminal repeat by UV-B irradiation requires a short sequence in the U3 region. *Arch Biochem Biophys* 1996; 326(2): 219-226.
- [79] Faure E, Best-Belpomme M, Champion S. UVB irradiation upregulation of the *Drosophila* 1731 retrotransposon LTR requires the same short sequence of U3 region in a human epithelial cell line as in *Drosophila* cells. *Photochem Photobiol* 1996; 64(5): 807-813.
- [80] Faure E, Best-Belpomme M, Champion S. X-irradiation activates the *Drosophila* 1731 retrotransposon LTR and stimulates secretion of an extracellular factor that induces the 1731-LTR transcription in nonirradiated cells. *J Biochem* 1996; 120(2): 313-319.
- [81] Faure E, Emanoil-Ravier R, Champion S. UVB irradiation-induced transcription from the long terminal repeat of intracisternal A particles and UVB-induced secretion of an extracellular factor that induces transcription of the intracisternal A particles in unirradiated cells. *J Photochem Photobiol B* 1996; 36(1): 61-66.
- [82] Faure E, Emanoil-Ravier R, Champion S. Induction of transcription from the long terminal repeat of the intracisternal particles type A

- (IAP) by X-irradiation. *Arch Physiol Biochem* 1997; 105(2): 183-189.
- [83] Wessler SR. Turned on by stress. *Plant retrotransposons*. *Curr Biol* 1996; 6(9): 959-961.
- [84] Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimate divergence. *Proc Natl Acad Sci USA* 2000; 97(12): 6603-6607.
- [85] Shim S, Lee SK, Han JK. A novel family of retrotransposons in *Xenopus* with a developmentally regulated expression. *Genesis* 2000; 26(3): 198-207.
- [86] Ramallo E, Kalendar R, Schulman AH, Martínez-Izquierdo JA. Remel, a Copia retrotransposon in melon, is transcriptionally induced by UV light. *Plant Mol Biol* 2008; 66(1-2): 137-150.
- [87] Faure E, Lecine P, Lipcey C, Champion S, Imbert J. Cell-to-cell contact activates the long terminal repeat of human immunodeficiency virus 1 through its kappaB motif. *Eur J Biochem* 1997; 244(2): 568-574.
- [88] Faure E, Cavard C, Zider A, Guillet JP, Resbeut M, Champion S. X irradiation-induced transcription from the HIV type 1 long terminal repeat. *AIDS Res Hum Retroviruses* 1995; 11(1): 41-43.
- [89] Faure E, Lecine P, Imbert J, Champion S. Activation of the HIV type 1 long terminal repeat by X-irradiation involves two main Rel/NF-kappa B DNA-binding complexes. *AIDS Res Hum Retroviruses* 1996; 12(16): 1519-1527.
- [90] Faure E, Lecine P, Imbert J, Champion S. Activation of the transcription from the human immunodeficiency virus type 1 (HIV-1) long terminal repeat by autologous and heterologous cell-to-cell contact. *Cell Mol Biol (Noisy-le-grand)* 1996; 42(6): 811-823.
- [91] Faure E, Rameil P, Lecine P, *et al.* Secretion of extracellular factor(s) induced by X-irradiation activates the HIV type 1 long terminal repeat through its kappaB motif. *AIDS Res Hum Retroviruses* 1998; 14(4): 353-365.
- [92] Kumar S, Orsini MJ, Lee JC, McDonnell PC, Debouck C, Young PR. Activation of the HIV-1 long terminal repeat by cytokines and environmental stress requires an active CSBP/p38 MAP kinase. *J Biol Chem* 1996; 271(48): 30864-30869.
- [93] Morales JF, Snow ET, Murnane JP. Environmental factors affecting transcription of the human L1 retrotransposon. II. Stressors. *Mutagenesis* 2003; 18(2): 151-158.
- [94] Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien MA. The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Mol Biol* 1997; 33(2): 257-266.
- [95] Takeda S, Sugimoto K, Otsuki H, Hirochika H. Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate. *Plant Mol Biol* 1998; 36(3): 365-376.
- [96] Grandbastien MA. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* 1998; 3(5): 181-187.
- [97] Capy P, Gasperi G, Biémont C, Bazin C. Stress and transposable elements: co-evolution or useful parasites? *Heredity* 2000; 85(Pt2): 101-106.
- [98] Nellåker C, Yao Y, Jones-Brando L, Mallet F, Yolken RH, Karlsson H. Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology* [serial on the Internet]. Jul 2006; [cited 2006 July 6] 3: 34 [about 16 screens]. Available from: <http://www.retrovirology.com/content/3/1/44>
- [99] Øresland V, Bray RA. Parasites and headless chaetognaths in the Indian Ocean. *Mar Biol* 2005; 147(3): 725-734.
- [100] Daponte MC, Gil de Pertierra AA, Palmieri MA, Ostrowski de Nuñez M. Monthly occurrence of parasites of the chaetognath *Sagitta friderici* off Mar del Plata, Argentina. *J Plankt Res*, in press 2008.
- [101] Nair S, Simidu U. Distribution and significance of heterotrophic marine bacteria with antibacterial activity. *Appl Environ Microbiol* 1987; 53(12): 2957-2962.
- [102] Duvert M, Perez Y, Casanova J-P. Wound healing and survival of beheaded chaetognaths. *J Mar Biol Assoc UK* 2000; 80(5): 891-898.

Received: March 28, 2008

Revised: April 8, 2008

Accepted: April 9, 2008

© Barthélémy *et al.*; Licensee Bentham Open.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.5/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.