BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(6)

Ty1-*copia* group retrotransposons and the evolution of retroelements in several angiosperm plants: evidence of horizontal transmission

Pasqualina Woodrow*, Loredana F Ciarmiello, Stefania Fantaccione, Maria Grazia Annunziata, Giovanni Pontecorvo & Petronia Carillo

II University of Naples, Department of Life Science, via Vivaldi 43, 81100 Caserta, Italy; Pasqualina Woodrow – Email: pasqualina woodrow@unina2.it; Phone: +39-0823-274542; Fax: +39-0823-274571; *Corresponding author

Received February 20, 2012; Accepted March 28, 2012; Published March 31, 2012

Abstract:
The phylogenetic relationships among thirty-seven new Ty1-*copia* group retrotransposons in seven angiosperm plants were examined by reverse transcriptase and ribonuclease H sequence analysis. Distribution pattern of the retrotransposons of closely related plant species generally reflects a close phylogenetic relationship. In contrast, we found that several retrotransposon sequences from the same genome exhibited a high degree of divergence and had a relatively high degree of identity versus retrotransposon sequences from widely divergent species, including an ancestral phytopathogen fungus. This finding supports the hypothesis that the horizontal transmission from phytopatogen organism to the host flowering plants could have played a role in the evolutionary dynamics of Ty1-*copia* group retrotransposons.

Key words: evolutionary dynamics, flowering plants, horizontal gene transfer, retrotransposons
Note: Nucleotide sequence data reported are available in the GenBank, accession nos: [DQ644578-DQ644614].

Background:
The retrotransposons constitute the most common class of repetitive DNA in a broad range of *taxa*, interspersed within the genome and ubiquitous in eukaryotes. It is currently believed that they have played a role in shaping genome architectures contributing significantly to the remarkable variations in genome size and evolution by changing the structures and expression patterns of genes **[1].** The advent of PCR methods for amplifying conserved domains of retrotranspons has led to the rapid increase of finding on sequence evolution and phylogenetic relationships of the Ty1-*copia* in plant genomes **[2, 3].** In this scenario, the prevailing view is that the Ty1-*copia* retrotransposons existed in the early phase of eukaryote life and diverged into heterogeneous subgroups before modern plant species aroused **[4].** Previous studies have shown large heterogeneity within LTR retroelements as a result of mutations accumulated during evolution. These mutations originated different lineages of retrotransposons, characterized by differences in activity and chromosomes distribution **[5].** Retrotransposon are usually located in intergenics regions and tend to accumulate in centromeres, telomeres and heterochromatic regions. In many cases, the same retrotransposon is found in different cultivars of the same species and in several related species, but not in far distant *taxa* **[2, 3].**

Phylogenetic analysis of reverse transcriptase (RT) sequences from the Ty1-*copia* retrotransposons showed that the sequence heterogeneity is generally proportional to the evolutionary distance of their host plant species, implying that the vertical

transmission has been a major factor in the evolution of Ty1-*copia* retrotransposons within the plants **[6].** On the other hand, based on the high sequence conservation across large evolutionary distances and a lower sequence similarity between elements from the same genome and related species, it has been postulated that the horizontal transmission (HT) of Ty1-*copia* retrotransposons can occur in plants **[3].** Few cases of HT involving transposable elements have been reported, and most involved Class II transposable elements **[7].** In this study we have isolated and characterized new Ty1-*copia* retrotransposons in seven dicotyledon plants and carried out a phylogenetic analysis. Distinct evolutionary lineages of Ty1-*like* were identified, and according to several authors significant variation among T*y* clade elements were observed. The similarity between *Solanum melongena* and *Phytophthora infestans* retroelements shown by molecular analysis and phylogenetic relationships was intriguing since these two species are unrelated. This similarity suggested that HT of the LTR retrotransposon named *Phythophthora* (GenBank accession no. CAI72292.1) has occurred within the *Solanum melongena* genome.

Methodology:

Plant material and genomic DNA isolation

Taxonomic classification of the studied angiosperm species is listed in **Table 1 (see supplemental material)** Plants were obtained from a nursery in South-Italy (Caserta). Total DNA from each plant species was extracted from the fresh leaf tissues according to Fantaccione and coworkers **[8].**

PCR strategy, internal controls, cloning and sequencing

Total DNA (2 µg) was completely digested with *Eco*RI (Roche) and digested samples (200 ng) were used for PCR amplifications with three combinations of degenerate oligonucleotide primers (RT1/RNase H1; RT2/RNase H1; RT3/RNase H1). The sense RT1 (5′-GATGTDAARACKRCNTTYYTD-3′), RT2 (5′-ATGGAGCARCCDGAMGGHTTY-3′), RT3 (5′-TATGTDGATGAYATGYTDATT-3′) and the antisense RNase H1 (5′-CCTCACATCWATRTGYTTBGW-3′) degenerate oligonucleotide primers, which respectively correspond to the conserved RT (DVKTA(T)FL, MEQPE(D)GF, YVDDMLI) and ribonuclease H (RNase H) (T(S)KHIDVR) peptide motifs of the Ty1-*copia* group retrotransposons were used. For each species, PCR amplifications (0.8-1 kb) were performed in 50 µL volume of 2 mM $MgCl_2$, 200 µM of each dNTP, 200 µM of each primer and 2 U of FastStart Taq DNA Polymerase (Roche). After an initial denaturation step for 4 min at 94°C, the PCR reactions were subjected to 30 cycles of amplification consisting of 1 min denaturation at 94°C, 1 min annealing at 44°C and 1 min extension at 72°C with a 10 min final extension at 72°C. PCR products were purified from the agarose gel, cloned in pDrive vector (Qiagen) and sequenced using ABI 377 automated DNA sequencer (Applied Biosystems). Internal PCR controls: *Pinci*1-26 forward (5′-GTATGTGGGGCTCCAAGCC-3′) and *Pinci*1-26 reverse (5′-GCGGCTGCACGGTCATTG-3′) were designed on *Pinci*1-26 noncoding infection-specific mRNA sequence of *P. infestans* (GenBank accession no. EF091740). The control PCR was carried out as described above. All clones were named as follows: The first letter meant cloning product, the second and third letter indicated the genus and species (p*C.a.*= *C. Annuum*, p*C.m.*= *C. Melo*, p*C.s.*= *C. Scolymus*, p*H.m.*= *H. Macrophylla*, p*P.a.*= *P. Avium*, p*S.m.*= *S. Melongena*, p*V.v.*= *V. Vinifera*) the fourth letter identified the numeric sequence of the screened clones.

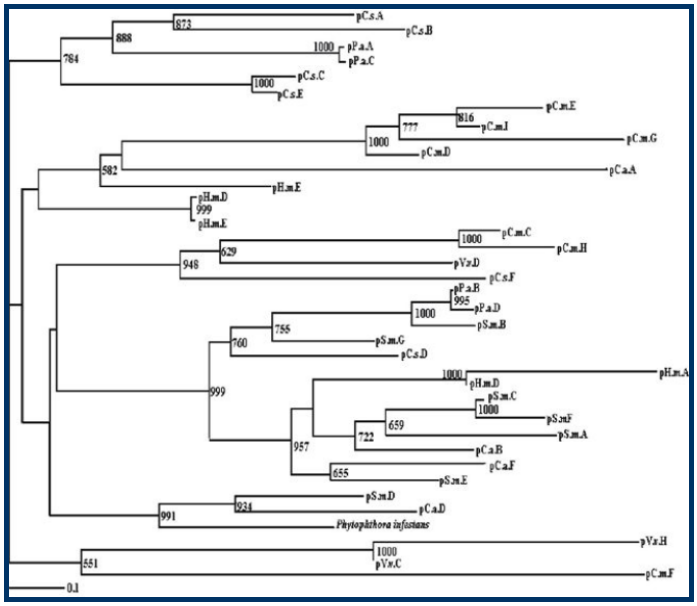

Figure 1: Phylogenetic evidence of horizontal transmission from *P. infestans* to their host flowering plants. Bootstrap values of 50% and higher are shown (1,000 trials).

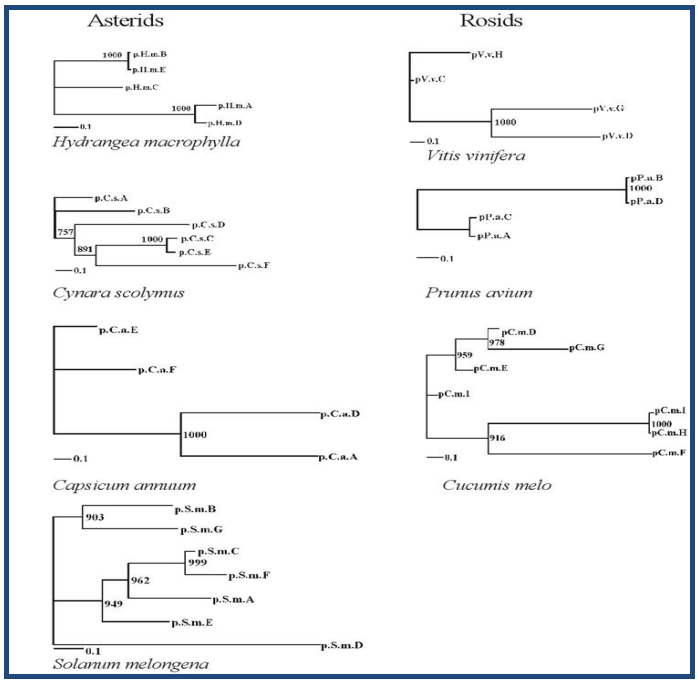

Figure 2: Phylogenetic trees of the Ty1-*copia* group retrotransposons from seven plant species based on RT and RNase H domains. Numbers on the branches are the bootstrap percentages for 1,000 replicates.

Databases, sequence analysis and phylogenetic trees

Nucleotide sequences were compared to the GenBank-NCBI database using the BLAST network service (http://www.ncbi.nlm.nih.gov/BLAST/) and analyzed using CENSOR software (http://www.girinst.org/censor). Multiple sequence alignments were performed utilizing the Clustal W

(1.8) program **[8]** from the DDBJ Homology Search system (http://www.ddbj.nig.ac.jp) with blosum matrix. The phylogenetic trees were constructed by Neighbor-Joining method **[9]** and plotted by DrawTree using the PHYLIP software package **[10].** Nonparametric bootstrap resampling (1000 replicates) was used to estimate the clade robustness. C-values of genome plants were estimated using the Angiosperm DNA C-values database available at http://www.rbgkew.org.uk./cval/database1.thml site.

Analysis of synonymous and nonsynonymous substitution per site and codon bias

p*C.a*.D and p*S.m*.D retrotransposons were compared to the *P. infestans* RT-ORF sequence (GenBank accession no. CAI72292.1) to examine the differences in synonymous and nonsynonymous changes using the SNAP (Synonymous/Nonsynonymous Analysis Program) tool in the HIV Sequence Database (http://www.hiv.lanl.gov). This program uses the algorithm devised by Nei and Gojobori **[11].** Codon bias as determined by the Nc value was computed using the Codon W program (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html).

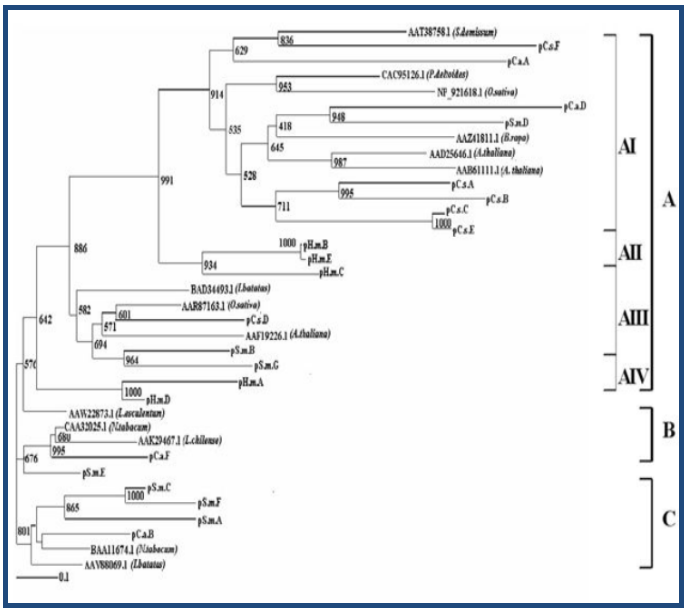

Figure 3: A phylogenetic tree based on the RT and RNase H domain analysis from Ty1-*copia* group retrotransposons in Asterids species and in different plant using the Neighbor-Joining method. Numerals adjacent to branches indicate are the bootstrap values for 1,000 replicates. Ty1-*copia* retrotransposons previously identified are shown as accession number and source species.

Discussion:

Identification of Ty1-copia retrotranspons in various plant species

In order to study the evolution of Ty1-*copia*, we performed a PCR-based assay using genomic DNAs from seven plant species. By using the degenerate oligonucleotide primers described in methodology, single products of the expected size (approx. 1000 bp) were obtained and 52 clones were sequenced **Table 1 (see supplemental material).** Based on a search against the GenBank database and CENSOR software, 37 sequences showed clear similarity (40%-60%) to RT-RnaseH domains **Table 1 (see supplemental material).** The majority of detected Ty1-*copia* elements revealed defective translated products for the presence of stop codons and short indels that resulted in frameshift and frame stop mutations, resulting transcriptionally or transpositionally inactive. The variations within the inactive elements can contribute to the evolution of the plant genome, leading to gene duplication events and providing new properties in the retrotransposition mechanisms **[12].**

Phylogenetic analysis of Ty1-copia-like retrotransposons in plants

Phylogenetic relationships were determined by comparing the thirty-seven clones among them or with retrotransposons from other plant species **(Figures 1-4 & Table 1, 2 see supplemental material)** Phylogenetic analysis in **(Figure 3)** showed that the retrotransposons were clustered in three clades (A, B and C). Clade A was further divided in four subclades (AI, AII, AIII and AIV), as judged by the bootstrap values. The subclades AI and AIII contained members from monocotyledon and dicotyledon plants. These results could be explained by the universal idea that the different retrotransposons already existed before the divergence between monocotyledon and dicotyledon plants and have been vertically transferred **[3].** Vertical relationships were found also among several *copia*-like retrotransposons from related plant species belonging to the Solanaceae family (*L. esculentum*, *N. tabacum*, *L. chilense*, *C. annuum*, *S. melongena* and *I. batatas*), as shown in subclades B and C **(Figure 3)**. However, vertical and horizontal transmissions are not mutually exclusive **[9].** In fact, p*S.m*.D and p*C.a*.D retrotransposons were clustered in the subclass AI together with same dicotyledon plant retrotransposons (*B. rapa* and *A. thaliana*). In addition, p*C.s*.D retrotransposon and a monocotyledon retrotransposon from *O. sativa* genome clustered together in the subclade AIII (79% of positivity score). Similar phylogenetic relationships were observed about plant retrotransposons belonging to the Rosids subclass, too **(Figure 4).**

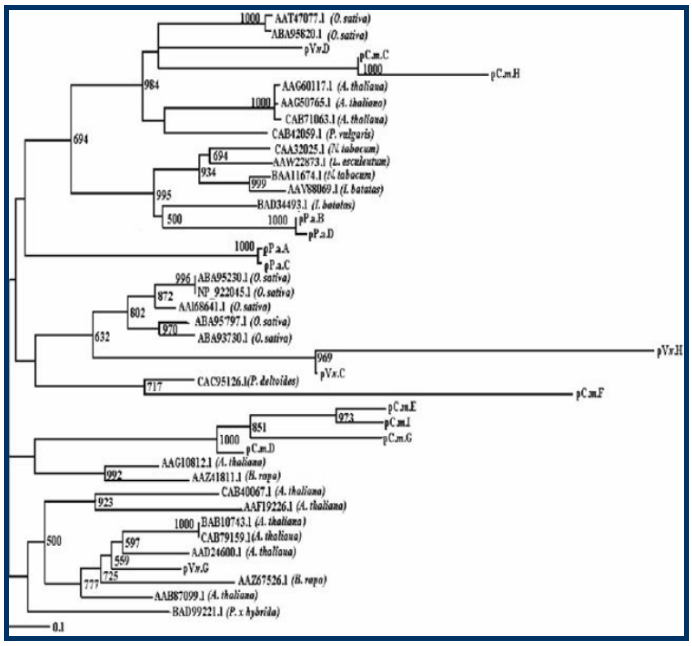

Figure 4: A phylogenetic tree based on the RT and RNase H domain analysis from Ty1-*copia* group retrotransposons in Rosids species and in different plant using the Neighbor-Joining method. Bootstrap values of 50% and higher are shown (1,000

trials). Ty1-*copia* retrotransposons previously identified are shown as accession number and source species.

Interestingly, when we submitted the p*C.a*.D and p*S.m*.D retrotransposons from *C. annuum* and *S. melongena* genomes (Asterids subclass and Solanaceae family) to a Blastx search (http://www.ncbi.nlm.nih.gov/BLAST/), the *P. infestans* fungal hit (GenBank accession no.CAI72292.1) matched with 59% and 63% positivity scores, respectively. This plant pathogen parasitizes a large number of host species, including *S. melongena* and *C. annuum* **[13, 14].** Although fungi and angiosperm plants last shared a common evolutionary ancestor about 2.5 billion years, p*C.a*.D and p*S.m*.D elements are included in the same group with *P. infestans* retrotransposon **(Figure 1).** Codon bias is a possible selective constraint on synonymous nucleotides **[15].** A bias in codon usage occurs when synonymous codons are not all used at the same frequency in coding DNA and such bias in codon usage could result from mutational pressure or from selective pressure **[16].** A common measure of codon bias is the effective number of codons (Nc), which can vary from 21 (when only one codon is effectively used for each amino acid) and 61 (when codons are used randomly) **[17].** The *P. infestans*, *S. melongena* and *C. annuum* retrotransposon sequences have an Nc value of 52.81, 51.84 and 47.60, respectively, consistent with only relatively moderate codon bias. Since selection operates much more efficiently on nonsynonymous base substitution **[17],** a common measure of selection for function is the ratio of synonymous to nonsynonymous substitution per site (dS/dN). When RT sequences from p*C.a*.D and p*S.m*.D retrotransposons were separately compared to *P. infestans* putative polyprotein using the SNAP program, the corrected frequencies of nonsynonymous substitutions for this region were 0.3408 and 0.3481, and the corrected frequencies of synonymous substitutions ranges were 0.8415 and 0.9013. Importantly, the ratio dS/dN for these two analysis pairs was only 2.4695 and 2.5895, consistent with a weak selection for function at the amino acid level. To eliminate any possible DNA contamination from host-oomycete interaction, we instituted internal PCR controls using primers designed on *Pinci*1-26 pathogen sequence from *P. Infestans* **[13]** and we found no PCR amplification products that are no contamination.

Conclusion:

Our analysis showed that the phylogenetic relationships between Ty1-*copia* retrotransposons in several plant genomes are complex and consistent with the universal idea that today the different retrotransposons pervade the entire plant kingdom as large and highly heterogeneous populations. The high degree of similarity between retrotransposons from divergent species and, in addition, the sequence divergence in the same species and/or related species suggests that the horizontal transmission events occurred in Ty1-*copia* evolution. In this scenario, we have attempted to understand whether the relatively high degree of the sequence conservation between the retrotransposons from pathogen oomycete *P. infestans* and two Asterids species may be a new example of horizontal gene transfer.

Taken together, our data support the horizontal transmission hypothesis since it seems unlikely that this high degree of the sequence conservation over such a length of time is the product of standard vertical transmission. Our example is a minor contribution and further studies are needed to exhaustively identify retrotransposons from plant species to gain further insight into the evolutionary history of the relationship between retrotransposons and their genome hosts.

Acknowledgements:

The authors are grateful to Joseph Sepe (II University of Naples, Caserta, Italy), for improving the manuscript.

References:

[1] Bennetzen JL, *Trends Microbiol*. 1996 **4**: 347 [PMID: 8885169]
[2] Flavell AJ *et al*. *Mol Gen Genet*. 1992 **231**: 233 [PMID: 1370976]
[3] Flavell AJ *et al*. *Genetica* 1997 **100:** 185 [PMID: 9440272]
[4] Stuart-Rogers C & Flavell AJ, *Mol Biol Evol*. 2001 **18:** 155 [PMID: 11158374]
[5] Natali L *et al*. *Genome*. 2006 **49**: 64 [PMID: 16462902]
[6] Hirochika H & Hirochika R. *Jpn J Genet*. 1993 **68**: 35 [PMID: 8391285]
[7] Roulin A *et al*. *Plant J*. 2008 **53**: 950 [PMID: 18088314]
[8] Fantaccione S *et al*. *Bioinformation* 2008 **2**: 311 [PMID: 18478086]
[9] Thompson JD *et al*. *Nuclecic Acids Res*. 1994 **22**: 4673 [PMID: 7984417]
[10] Saitou N & Nei M, *Mol Biol Evol*. 1987 **4:** 406 [PMID: 3447015]
[11] Felsenstein J, *Meth Enzymol*. 1996 **266**: 418 [PMID: 8743697]
[12] Woodrow P *et al*. *Theor Appl Gen*. 2010 **121**: 311 [PMID: 20237753]
[13] Gbadegesin MA *et al*. *Mol Genet Genomics*. 2008 **280**: 305 [PMID: 18636276]
[14] Avrova AO *et al*. *Microbiol*. 2007 **153**: 747 [PMID: 17322195]
[15] Whittle CA *et al*. *BMC Genom*. 2007 **8**: 169 [PMID: 17567911]
[16] Wright F, *Gene* 1990 **87**: 23 [PMID: 2110097]
[17] Nei M & Gojobori T, *Mol Biol Evol*. 1986 **3**: 418 [PMID: 3444411]

Edited by P Kangueane

Citation: Woodrow *et al*. Bioinformation 8(6): 267-271 (2012)

BIOINFORMATION

open access

Supplementry material:

Table 1: Plant Species that were tested by PCR Assay

Class	Subclass	Order	Family	Species	PCR	Number of Sequenced Clones[a]
Magnoliopsida	Asterids	Cornales	Hydrangeaceae	*H. macrophylla (Thunb.) Ser*	+	5
		Asterales	Asteraceae	*C. scolymus L.*	+	6
		Solanares	Solanaceae	*C. annum L.*	+	4
				S.Melongena	+	7
	Rosids	Rosids incertae sedis	Vitaceae	*V. vinifera L.*	+	4
		Rosales	Rosaceae	*P. avium (L.) L.*	+	4
		Cucurbitales	Cucurbitaceae	*C. melo L.*	+	7

A Ty1-*copia* group retrotransposons isolated in this work

Table 2: Source of DNA sequences used in the phylogenetic analysis

Division (Class)	Species	Database Locus Name (element name)	Accession number [a]
Anthophyta (Monocotyledopsida)	*Oriza sativa*	NM_196636	NP_921618
		AC135792	AAR87163.1
		AC121365	AAT47077.1
		ABA95820	ABA95820.1
		AF458765(0st3-1)	AAL68641.1
		ABA95230	ABA95230.1
		NM_197063.1	NP_922045.1
		ABA95797	ABA95797.1
		ABA93730	ABA93730.1
(Dicotyledopsida)	*Populus deltoids*	PDE416708 (Retrotrop)	CAC95126.1
	Brassica rapa	AC166739 (01p13-1)	AAZ41811.1
	Phaseolus vulgaris	AC155340 (Br5)	AAZ67526.1
	Arabidopsis thaliana	PVA005761 (Tpv2-1C)	CAB42059.1
		AC007109	AAD25646.1
		F20P5	AAB61111.1
		AC007505 (Ta1-3-like)	AAF19226.1
		AC073555	AAG60117.1
		AC079131.4	AAG50765.1
		AAT20K12	CAB71063.1
		ATF25I24	CAB0067.1
		AC018460	AAG10812.1
		AC005825	AAD24600.1
		AB010695	BAB10743.1
		ATCHRIV56	CAB79159.1
		AC002391	AAB8799.1
	Solanum demissum	AC149291	AAT38758
	Nicotiana tabacum	NTTNTT194 (Tnt1)	CAA32025.1
		TOBAA (Tto1)	BAA11674.1
	Lycopersicon esculentum	AY6782298 (Silvia)	AAW22873.1
	Ipomoea batatas	AY830138	AAV88069.1
		AB162659 (Rtsp-1)	BAD34493.1
	Lycopersicon chilense	AF279585 (TLC1.1)	AAK29467.1
	Solanum tuberosum	AY730334	AAU89728.1
		AY730334	AAU89730.1
	Petunia hybryda	AB196430 (rTph1)	BAD99221.1