Volume 319, number 3, 261–266 FEBS 12247 March 1993

Reverse transcriptase families and a *copia*-like retrotransposon, *Osser*, in the green alga *Volvox carteri*

Andreas Lindauer, David Fraser, Martina Brüderlein and Rüdiger Schmitt

Lehrstuhl für Genetik, Universität Regensburg, D-8400 Regensburg, Germany

Received 14 December 1992; revised version received 18 January 1993

By using the polymerase chain reaction (PCR) we have isolated and sequenced two distinct families of reverse transcriptase (RT) sequences from the genome of the colonial alga, *Volvox carteri*. Probing a genomic library with these RT clones revealed *copia*-like retrotransposons. One of these elements, named *Osser*, is 4,875 bp long, bordered by 197-bp identical long terminal repeats (LTRs), and shows the typical organization of retrotransposons belonging to the *copia-Ty1* group. This is the first complete *copia*-like retrotransposon sequence described in a green alga.

Reverse transcriptase; Retrotransposon; *copia-Ty1* element; Green alga; *Volvox carteri*

1. INTRODUCTION

Retroelements, including the vertebrate retroviruses [1] and the long terminal repeat (LTR) retrotransposons of *Drosophila* [2] and *Saccharomyces cerevisiae* [3], are among the most common and widespread types of eukaryotic transposable elements [4]. The non-viral retroelements can be grouped into (i) the retrotransposons, (ii) the retroposons, (iii) the retrons (in bacteria) and (iv) the retrosequence group [5]. Retrotransposons transpose by means of an RNA intermediate by a mechanism equivalent to that of retroviruses. They are usually divided into two groups corresponding to the best-known elements: the *gypsy-Ty3* group and the *copia-Ty1* group (here referred to as *copia*-like elements), which differ in the linear arrangement of the enzymatic functions encoded by the *pol* gene [6]. Retrotransposons can be distinguished from the retroposons (non-LTR retrotransposons) by the presence of LTRs. These LTRs are variable in size, carry the signals for transcription initiation and termination, and flank an internal domain encoding proteins analogous to the *gag* and *pol* of retroviruses. The *pol* gene encodes several enzymatic functions, including protease, integrase, reverse transcriptase (RT) and RNase H [1]. The most highly conserved part of the *pol* gene is the reverse transcriptase portion [4].

Recent reports [7–9] have revealed that *copia*-like retrotransposons are ubiquitous components of plant genomes; but only some *copia*-like retrotransposons, like *Ta1* [10], *Tnt1* [11], *Tst1* [12], *Tos3-1* [13] and *Wis-2* [14], of higher plants have been fully characterized. Voytas et al. [7] previously demonstrated the amplification of two partial RT sequences in *Volvox carteri* belonging to elements of the *copia-Ty1* group. So far the only completely described retroelements in green alga are the *TOC1* elements [15] of *Chlamydomonas reinhardtii*, which exhibit a quite unusual structure unlike that of any of the known retroelements. However, no complete *copia*-like algal retrotransposon has been reported prior to this communication. In higher plants, transposition has been demonstrated only for *Tnt1* [11] and *Bs1*, however, these were restricted to unusual stress situations [16], what McClintock called "genomic shocks" [17]. The stress-induced activity of retroelements is thought to facilitate genomic restructuring and, hence, adaptation to extreme environmental conditions [17]. Mobile retrotransposons also provide potential tools for plant genetic analysis by gene tagging [18].

It has been shown previously that a pair of degenerate oligonucleotide primers based on two highly conserved domains of RT can be successfully used for PCR amplification of *copia*-like sequences in *Arabidopsis thaliana* [19]. By using the same strategy, we have so far identified and analyzed two families of *copia*-like RT sequences with eight and five members, respectively, in the genome of the green alga *V. carteri*. By probing a genomic library with these cloned sequences we have isolated and characterized *copia-Ty1* group elements in

Correspondence address: R. Schmitt, Lehrstuhl für Genetik, Universität Regensburg, Universitätsstr. 31, D-8400 Regensburg, Germany. Fax: (49) (941) 943-3163.

Abbreviations. A., *Arabidopsis*; aa, amino acid(s); bp, base pair(s); kb, kilobase(s) or 1,000 bp; LTR, long terminal repeat; nt, nucleotide(s); RT, reverse transcriptase; TS, target-site duplication; VCRT, *Volvox copia*-like reverse transcriptase; *V.*, *Volvox*.

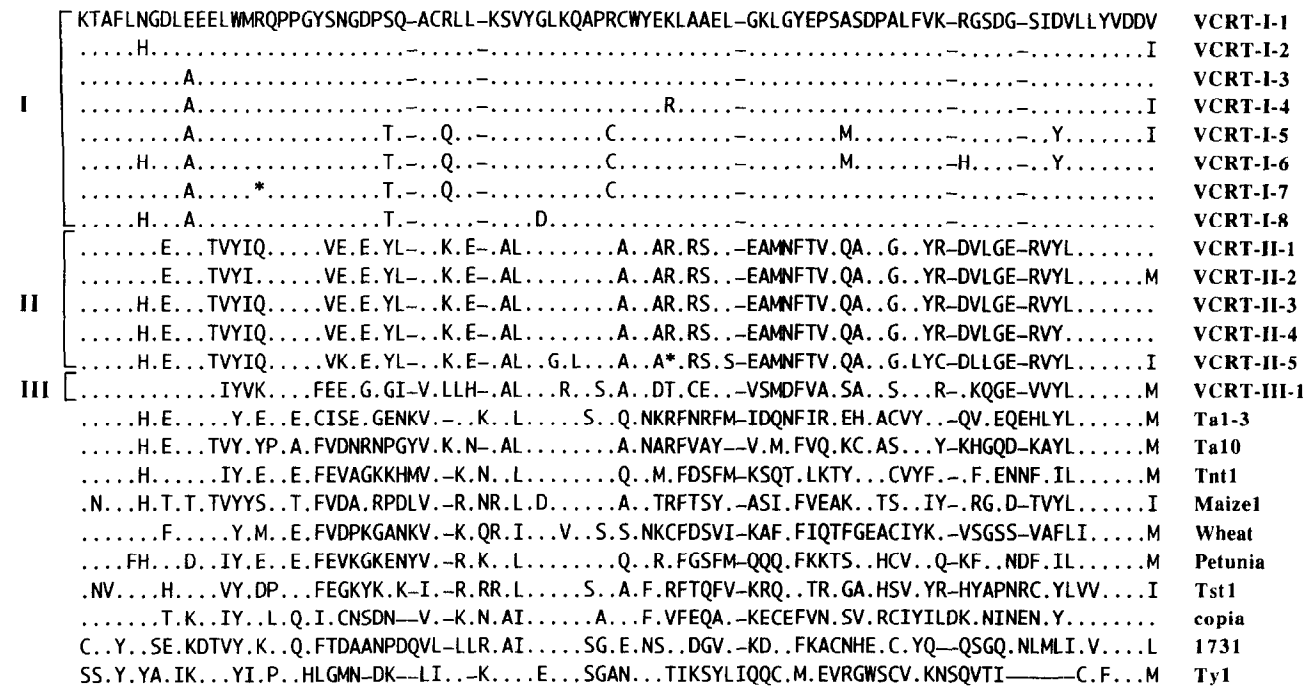

Fig. 1. Alignment of derived amino acid sequences of a RT region from 13 independent clones obtained by PCR amplification of *V. carteri* genomic DNA. Sequences fall into three families: VCRT-I represented by 8 different sequences, VCRT-II represented by 5 sequences and VCRT-III [7] represented by 1 sequence. Dots symbolize amino acids identical to VCRT-I-1 (top line). Asterisks denote nonsense codons and hyphens mark gaps due to the alignment. Corresponding peptide sequences derived from related *copia*-elements, like *Ta1-3* and *Ta10* (*A. thaliana*), *Tnt1* (*N. tabacum*), *Tst1* (*S. tuberosum*), *copia* and *1731* (*D. melanogaster*), *Ty1* (*S. cerevisiae*) and parts of the RT sequences of maize, petunia and wheat are included in the alignment.

Volvox. One of these, named *Osser*, has been completely sequenced and will be described here.

2. MATERIALS AND METHODS

V. carteri f. *nagariensis* strain HK10 was grown as previously described [20] and DNA was prepared using the standard procedure [20].

PCR was used to amplify and clone a portion of *copia*-like RT from *V. carteri* genomic DNA using two conserved primers. Oligonucleotide primers, synthesized in an Applied Biosystems 381A synthesizer, were based on the consensus sequence of the RT portion of the *pol* gene, as previously reported [19]: primer 1, 5′ AYRTCRTCNA-CRTANAGNAG-3′; primer 2, 5′ AARACNGCNTTYT-TRMAYGG-3′, where M = A + C, N = A + C + G + T, R = A + G and Y = T + C. PCR reactions were performed in 100 μl mixtures with 200 ng of genomic DNA, each dNTP at 200 μM, 200 pmol of each primer and buffer provided by the supplier of *Taq* polymerase (Perkin-Elmer/Cetus, Norwalk, CT, USA). Temperature cycling on a Perkin-Elmer TC1 thermocycler was performed with the following profile: 94°C for 1 min, 48°C for 1 min, and 72°C for 3 min through 30 cycles. After chloroform extraction of the reaction mixtures, the amplification products were precipitated and gel purified. They were then phosphorylated with T4 DNA polynucleotide kinase, blunt-ended with T4 DNA polymerase and cloned into the pUC8 derivative pJOE533 (J. Altenbuchner, pers. commun.) linearized with *Sma*I. DNA sequence data were obtained by dideoxy-sequencing [21] for 20 cloned sequences.

A λEMBL3-based genomic library of *V. carteri* [20] was screened by plaque hybridization using the probes VCRT-I-1 and VCRT-II-1 (Fig. 1). Plaques were transferred to Hybond N+ nylon filters (Amersham Buchler, Braunschweig, Germany), DNA probes were labeled by random priming (Boehringer-Mannheim, Germany) and hybridized to the filters at 58°C using the conditions recommended by the manufacturer. Filters were washed twice for 20 min in 2 × SSC/0.1% SDS at room temperature and subsequently in the same buffer at 50°C. Positive plaques were selected, purified twice and DNA was restriction analyzed. DNA was subcloned into pUCBM20 (Boehringer-Mannheim, Germany) vector and DNA sequences were obtained using the dideoxy method [21] with Sequenase (USB Corp., Cleveland, OH, USA) on double-stranded DNA templates. The GCG programs [24] were used for phylogenetic analysis and sequence comparisons on a VAX 3100 and on a Comparex (DKFZ Heidelberg, Germany) computer. The sequences will appear in the EMBL Nucleotide Sequence Database under accession numbers X69621–X69628 (VCRTI: 1 to 8), X69629–X69633 (VCRTII: 1 to 5) and X69552 (*Osser*).

3. RESULTS AND DISCUSSION

3.1. *PCR amplification of RT partial sequences*

This study was designed to use the PCR technique for detecting and isolating *copia*-like retroelements in the green alga *V. carteri*. The priming regions chosen specify the peptide sequences KTAFLH/NG and LLYVDDM/V [19], which are well-conserved between various *copia-Ty1* group members. By using deduced degenerate oligonucleotide primers that correspond to these highly conserved sequences in the RT region of the *pol* gene, we have amplified RT regions of presumptive retrotransposons from *V. carteri*. These PCR primers have previously been shown to amplify a part of the RT genes of *copia*-like retrotransposons of *A. thaliana* [19]

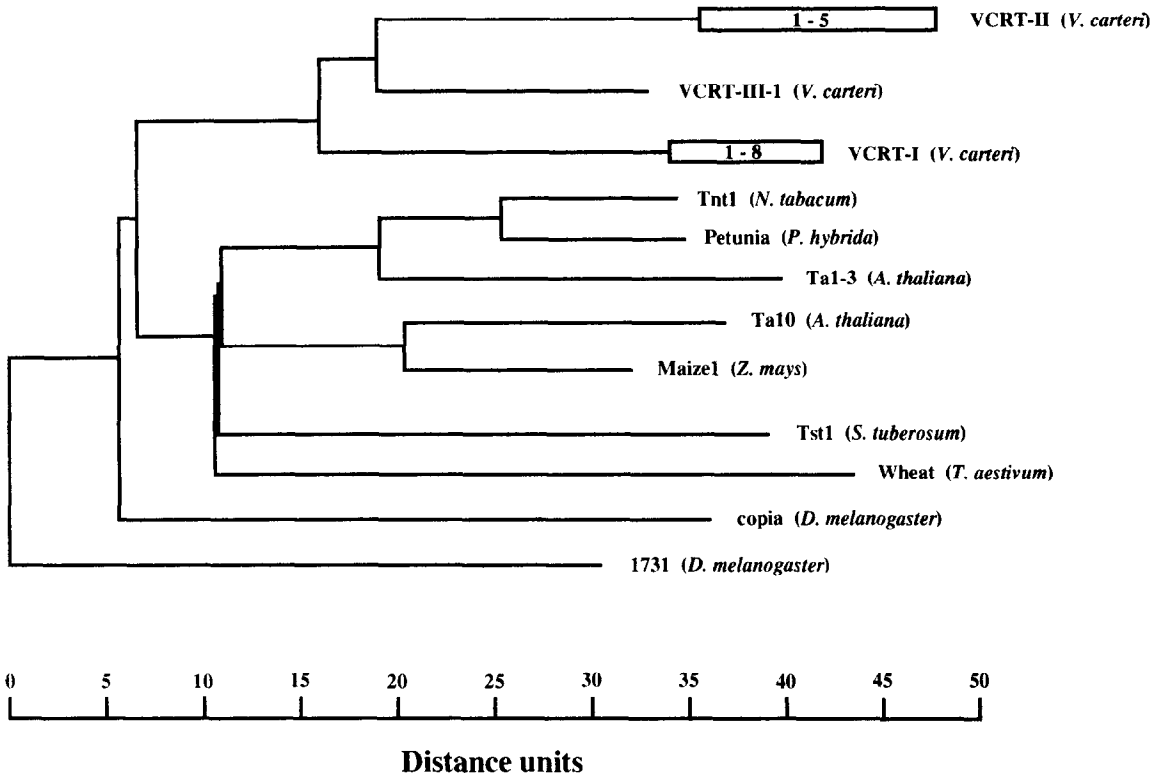

Fig. 2. Phylogenetic tree of *copia*-like RT sequences from Fig. 1. Divergences in distance units are indicated by horizontal branch lengths [29], whereas vertical lengths are without any significance. Boxes indicate minor divergences within the members of the families VCRT-I (1–8) and VCRT-II (1–5). The *Ty1* element was used to root the tree.

and of 64 other plant species [7]. The amplification products from *Volvox* DNA migrated as a major band in an 1.6% agarose gel with the expected size of about 260 bp. DNA fragments representing this major band were cloned, and a total of 20 clones were further analyzed by sequencing.

Comparisons of the deduced amino acid sequences of these clones revealed the existence of 13 variant peptide sequences (the other 7 clones being repetitions), which have been grouped into two distinct families (VCRT-I and VCRT-II) with eight and five members, respectively. Recently, Voytas et al. [7] presented two additional RT sequences of *V. carteri* named Volvox1 and Volvox2. According to our analysis, Volvox1 represents a third family (VCRT-III), whereas Volvox2 is identical to VCRT-II-2, described here. Fig. 1 shows the amino acid sequence alignments of all members of the VCRT (*Volvox copia*-like RT) families. These sequences are also compared to corresponding regions of other membranes of the *copia-Ty1* group like *Ta1-3* and *Ta10* [19], *Tnt1* [11], *Tst1* [12], *copia* [25], *1731* [26], *Ty1* [27] and PCR fragments of wheat, maize and petunia [7]. Using the GCG-TREE program [29], we generated a phylogenetic tree (Fig. 2) based on the above amino acid sequence alignments. The resulting diagram (Fig. 2) suggests that the three VCRT families share a common ancestor and form a monophyletic clade that is composed exclusively of algal retrotransposons. There appears to be a correlation between the degree of similarity between *copia*-like elements and the phylogenetic distances between species harboring these RT sequences suggesting that the inferred horizontal transmission of retrotransposons between different species [9] must have occurred early in the evolution of the various phyla.

3.2. *Cloning of retrotransposons and characterization of the element* Osser

Screening of a genomic library of *V. carteri* HK10 with a mixture (1:1) of VCRT-I-1 and VCRT-II-1 DNAs (Fig. 2) yielded ca. 400 positive plaques from five genome equivalents indicating a copy number of about 80 elements per genome. Eight clones (λVCRT1 to 8) were randomly selected and further purified. Hybridizing DNA fragments of λVCRT2 showing the strongest signal were subcloned and sequenced. The sequence analysis revealed that this clone bore a retroelement, which we have named *Osser*. The (sequence-derived) structure of *Osser* is similar to other retrotransposons, like *copia*, *Tnt1* and the *Ta1-3*. The 4,875-bp retroelement *Osser* is flanked by a 5-bp duplication of host DNA at the integration site. The large central domain is flanked by two identical, direct LTRs 197 bp in length (Fig. 3). These LTRs are terminated by short 6-bp in-

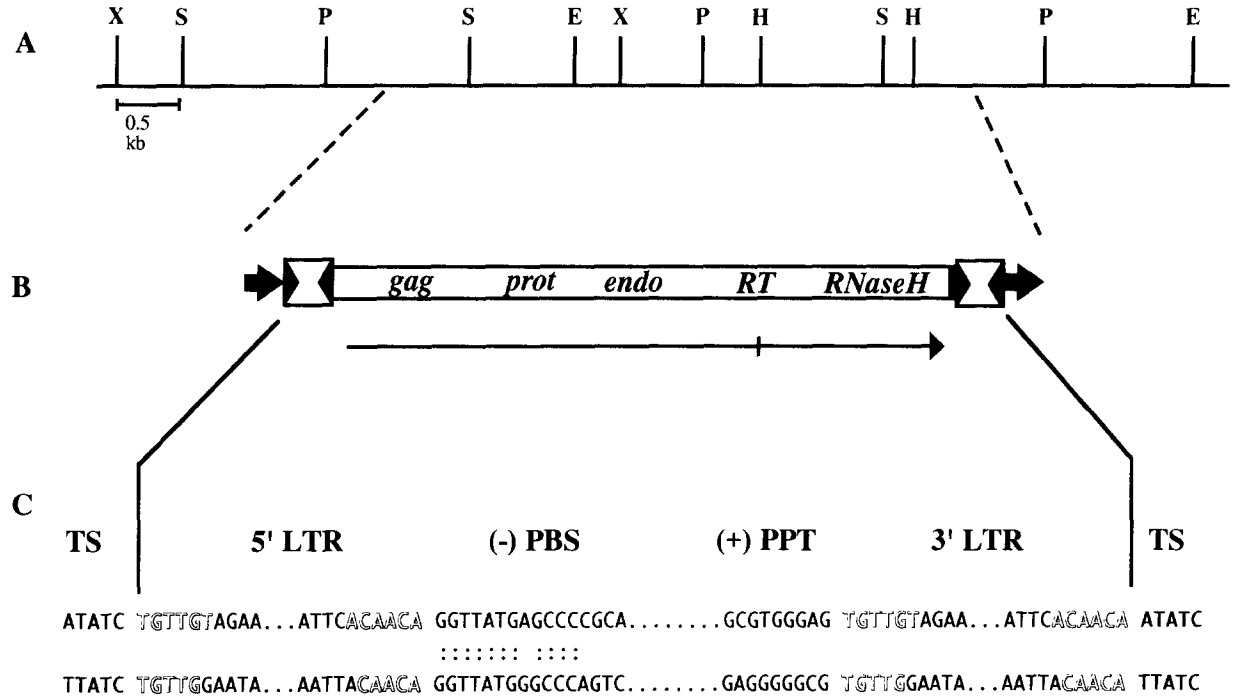

Fig. 3. Restriction map, overall organization and structural features of the retrotransposon *Osser* compared with sequence features of the *copia* element. (A) Restriction map of *Osser* with 5′ and 3′ flanking sequences. E = *Eco*RI, H = *Hin*cII, P = *Pst*I, S = *Sac*I and X = *Xba*I. (B) Organization of a *copia*-like retroelement with LTRs. Boxes with triangles mark the LTRs, which are bordered by 6-bp inverted repeats. *gag* = group-specific antigen, *prot* = protease, *endo* = endonuclease, *RT* = reverse transcriptase and *RNase H* = ribonuclease. 5′ and 3′ flanking bold arrows represent 5-bp direct repeats of the host target DNA generated upon insertion. The arrow below the central domain symbolizes the polyprotein-encoding ORF. The observed interruption of the RT domain by an 1-bp insertion, which may be overread by 'translational slippage' [33], is marked by a vertical line. (C) Special sequence features of *Osser* and *copia*, are highlighted by alignment Target site duplications (TS) are shown in bold, inverted repeats of the 5′ and 3′ LTRs by shadowing: identical nucleotides in the primer-binding sites (PBS) of *Osser* and *copia* are indicated by colons, (−) PBS, primer-binding site, complementary to the 3′ end of a host tRNA, is used for the synthesis of the (−) DNA strand; (+) PPT, polypurine tract, is used for the synthesis of the (+) DNA strand.

verted repeats, and the two terminal nucleotides of each LTR (5′ TG...CA 3′) are identical to those found in other retroviral-like elements [31]. The two identical LTR sequences of *Osser* with putative transcription signals are presented in Fig. 4. A potential TATA box is located at position 99, whereas the downstream initiation factor (DIF, [30]) binding site (5′ CGTG 3′) is at position 163. A polyadenylation signal (PAS), identical to the TGTAA signal found in *Volvox* nuclear genes [22,23] was detected at position 128. The sequence 5′ TTG 3′, 28 bp downstream from the polyA signal (RTS, Fig. 4), represents a sequence essential for the termination of viral RNA synthesis [31]. Moreover, these retroelements contain specific sequences at the boundaries of the central domain that serve to prime DNA synthesis by the RT. These primer sites (Fig. 3) consist of a short region of tRNA homology (*Osser*: $tRNA_i^{Met}$) at the 5′ end of the central domain (PBS) and a polypurine-rich sequence (PPT, polypurine tract) at its 3′ end. The PBS of *Osser* is nearly identical to the ones of *Ty5* [28] and *copia* [25], both of which possess an unconventional primer-binding site. Usually the 3′ OH of a cellular tRNA, which hybridizes to the PBS, is used to prime minus-strand DNA synthesis. However, in the case of *copia* it was demonstrated [32] that an internal portion of the $tRNA_i^{Met}$ is used as a primer. Therefore, like *Ty5* and *copia*, *Osser* differs from the plant elements, which

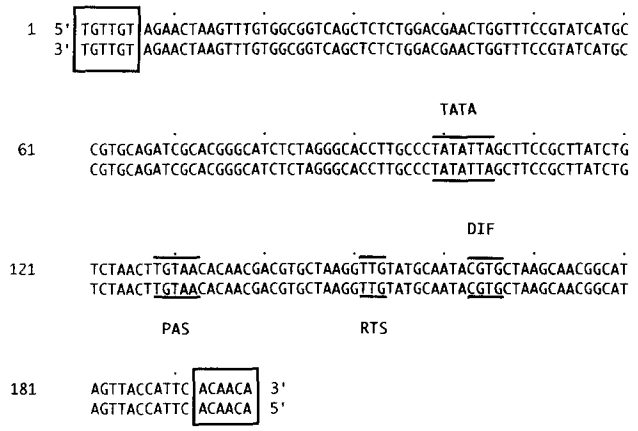

Fig. 4. Nucleotide sequence of 5′ and 3′ LTRs of *Osser*. Numbering refers to the 5′ LTR. Inverted repeats (6 bp) are boxed, and putative transcription initiation and termination signals are marked by lines. TATA, promoter box, and DIF, downstream inhibition signal (functional in the 5′ LTR); PAS, polyadenylation signal and RTS, RNA termination signal (functional in the 3′ LTR).

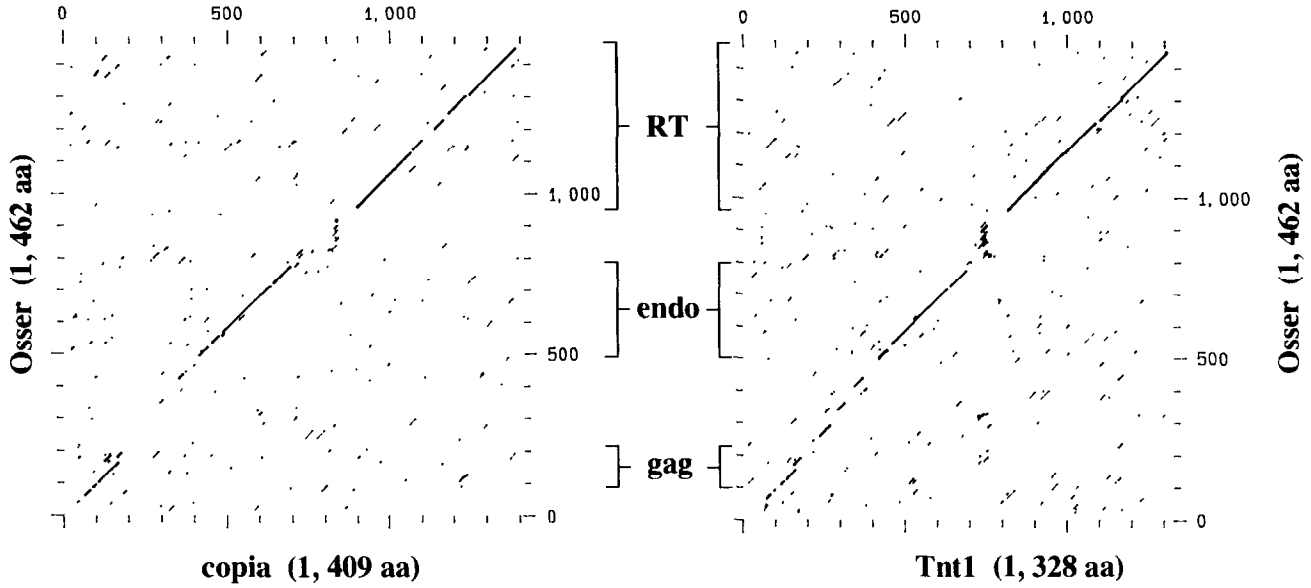

Fig. 5. Dot matrices for comparison of *Osser*-, *copia*- and *Tnt1*-derived polyprotein sequences. Matrices were computer generated using a window of 30 amino acids for comparison [24]. These revealed similarities between deduced polypeptides from *Osser* and *copia* (left; *gag* 43%; *endo* 56%; *RT* 56%) or *Osser* and *Tnt1* (right; *gag* 47%; *endo* 56%; *RT* 61%), respectively. Diagonal lines delineate regions of similarity marked by the approximate regions corresponding to the group antigen (gag), endonuclease (endo) and reverse transcriptase (RT) domains.

all use the 3′ OH of $tRNA_i^{Met}$ as a primer. The internal 4.4 kb region of *Osser* contains a single open reading frame (interrupted by a 1-bp insertion in the RT domain, deduced from amino acid sequence comparisons) that encodes a presumptive 1,462 amino acid polyprotein. Expression of a full-length polyprotein would require translational frameshifting, a mechanism operative in the synthesis of the gag–pol fusion protein of *HIV-1* [33]. But it cannot be excluded that it is simply a mutation. The central ORF of *Osser* is similar in structure to the ones of *copia* [25] and *Tnt1* [11], including their general organization as *gag* and *pol* genes, total length (*copia* 1,409 amino acid; *Tnt1* 1,328 amino acid), and the absence of the *env* domain of retroviruses. Identical to *copia*-like elements, the linear order of functional domains encoded by the *pol* gene of *Osser* is: protease, endonuclease, RT and RNase H. Dot matrix comparisons of derived amino acid sequences from *Osser* and *copia* and from *Osser* and *Tnt1* are presented in Fig. 5. In BESTFIT analyses the highest degree of identical amino acid residues were detected among the gag proteins (*copia* 21%; Tnt1 22%), the endonuclease domain (*copia* 33%; Tnt1 38%) and the RT domain (*copia* 36%; Tnt1 41%). With its RT sequence, *Osser* clearly belongs to the VCRT-II family (Fig. 1). The fact that the 5′ and 3′ LTRs are identical in sequence indicates that *Osser* has transposed in recent time. Other preliminary data suggest that *Osser* is still mobile in *Volvox* and may be used for gene tagging.

Thus, *Osser*, the first complete algal *copia*-like retroelement, encodes peptide sequences that resemble the conserved regions of the *copia-Ty1* group of retrotransposons. This expands and supports the previously postulated ubiquity of *copia*-like retroelements in all eukaryotic kingdoms and implies that this group of transposable elements shares a common ancestor. Moreover, the amazing sequence similarities among different species suggest that horizontal transfer of these elements may have ensued early-on [9].

Acknowledgements: We thank Andrzej Konieczny and Daniel Voytas for communicating pre-publication results and for sharing RT clones, and Steve Miller for critical review of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 43).

REFERENCES

[1] Varmus, H.E. and Brown, P. (1989) in: Mobile DNA (D.E. Berg and M.M. Howe, eds.) pp. 53–108, American Society for Microbiology, Washington, DC.
[2] Bingham, P.M. and Zachar, Z. (1989) in: Mobile DNA (D.E. Berg and M.M. Howe, eds.) pp. 485–502, American Society for Microbiology, Washington, DC.
[3] Boecke, J. (1989) in: Mobile DNA (D.E. Berg and M.M. Howe, eds.) pp. 335–374, American Society for Microbiology, Washington, DC.
[4] Doolittle, R.F., Feng, D.F., Johnson, M.S. and McClure, M.A. (1989) Quart. Rev. Biol. 64, 1–29.
[5] Hull, R. and Will, H. (1989) Trends Genet. 5, 357–359.
[6] Boeke, J.D. and Corces, V.G. (1989) Annu. Rev. Microbiol. 43, 403–434.
[7] Voytas, D.F., Cummings, M.P., Konieczny, A., Ausubel, F.M. and Rodermel, S.R. (1992) Proc. Natl. Acad. Sci. USA 89, 7124–7128.
[8] Flavell, A.J., Dunbar, E., Anderson, R., Pearce, S.R., Hartley, R. and Kumar, A. (1992) Nucleic Acids Res. 20, 3639–3644.
[9] Flavell, A.J., Smith, D.B. and Kumar, A. (1992) Mol. Gen. Genet. 231, 233–242.

[10] Voytas, D.F. and Ausubel, F.M. (1988) Nature 336, 242–244.
[11] Grandbastien, M.-A., Spielmann, A. and Caboche, M. (1989) Nature 337, 376–380.
[12] Camirand, A., St.-Pierre, B., Marineau, C. and Brisson, N. (1990) Mol. Gen. Genet. 224, 33–39.
[13] Hirochika, H., Fukuchi, A. and Kikuchi, F. (1992) Mol. Gen. Genet. 233, 209–216.
[14] Harberd, N.P., Flavell, R.B. and Thompson, R.D. (1987) Mol. Gen. Genet. 209, 326–332.
[15] Day, A., Schirmer-Rahire, M., Kuchka, M.R., Mayfield, S.P. and Rochaix, J.-D. (1988) EMBO J. 7, 1917–1927.
[16] Grandbastien, M.-A. (1992) Trends Genet. 8, 103–108.
[17] McClintock, B. (1984) Science 226, 792–801.
[18] Gierl, A. and Saedler, H. (1992) Plant Mol. Biol. 19, 39–49.
[19] Konieczny, A., Voytas, D.F., Cummings, M.P. and Ausubel, F.M. (1991) Genetics 127, 801–809.
[20] Mages, W., Salbaum, J.M., Harper, J.F. and Schmitt, R. (1988) Mol. Gen. Genet. 213, 449–458.
[21] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[22] Cresnar, B., Mages, W., Müller, K., Salbaum, J.M. and Schmitt, R. (1990) Curr. Genet. 18, 337–346.
[23] Schmitt, R., Fabry, S. and Kirk, D.L. (1992) Int. Rev. Cytol. 139, 189–265.
[24] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[25] Mount, S.M. and Rubin, G.M. (1985) Mol. Cell. Biol. 6, 1630–1638.
[26] Fourcade-Peronnet, F., D'Auriol, L., Becker, J., Galibert, F. and Best-Belpomme, M. (1988) Nucleic Acids Res. 16, 6113–6125.
[27] Clare, J. and Farabaugh, P. (1985) Proc. Natl. Acad. Sci. USA 82, 2829–2833.
[28] Voytas, D.F. and Boeke, J.D. (1992) Nature 358, 717.
[29] Feng, D.-F. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.
[30] Arkhipova, I.R. and Ilyin, Y.V. (1991) EMBO J. 10, 1169–1177.
[31] Christy, R.J., Brown, A.R., Gourlie, B.B. and Huang, R.C.C. (1985) Nucleic Acids Res. 13, 289–302.
[32] Kikuchi, Y., Yumiko, A. and Shiba, T. (1986) Nature 323, 824–826.
[33] Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988) Nature 331, 280–283.
[34] Temin, H.M. (1985) Mol. Biol. Evol. 2, 455–468.