Mol Gen Genet (1998) 259: 354–362

ORIGINAL PAPER

A. Hua-Van · F. Héricourt · P. Capy
M.-J. Daboussi · T. Langin

Three highly divergent subfamilies of the *impala* transposable element coexist in the genome of the fungus *Fusarium oxysporum*

Received: 2 December 1997 / Accepted: 28 April 1998

Abstract The transposable element *impala* is a member of the widespread superfamily of *Tc1-mariner* transposons, identified in the genome of the plant pathogenic fungus *Fusarium oxysporum*. This element is present in a low copy number and is actively transposed in the *F. oxysporum* strain F24 that is pathogenic for melons. The structure of the *impala* family was investigated by cloning and sequencing all the genomic copies. The analysis revealed that this family is composed of full-length and truncated copies. Four copies contained a long open reading frame that could potentially encode a transposase of 340 amino acids. The presence of conserved functional domains (a nuclear localisation signal, a catalytic DDE domain and a DNA-binding domain) suggests that these four copies may be autonomous elements. Sequence comparisons and phylogenetic analysis of the *impala* copies defined three subfamilies, which differ by a high level of nucleotide polymorphism (around 20%). The coexistence of these divergent subfamilies in the same genome may indicate that the *impala* family is of ancient origin and/or that it arose by successive horizontal transmission events.

Key words *Tc1-mariner* · Filamentous fungi · *Fusarium oxysporum* · Transposase · Evolution

A. Hua-Van (✉) · F. Héricourt[1] · T. Langin[2] · M.-J. Daboussi
Institut de Génétique et Microbiologie
Université Paris-Sud, 91405 Orsay Cedex, France
Fax: +33-1-6915-7006; e-mail: ahuavan@igmors.u-psud.fr

P. Capy
Laboratoire Populations, Génétique et Evolution
CNRS, 91198 Gif/Yvette Cedex, France

Present addresses:
[1] Institut Jacques Monod, 2 Place Jussieu
75251 Paris Cedex 05, France
[2] Institut de Biotechnologie des Plantes
Université Paris-Sud 91405 Orsay Cedex, France

Introduction

Fusarium oxysporum is a filamentous, soil-borne pathogenic fungus that attacks numerous plants. This fungus contains several families of transposable elements, representing the major classes of transposable elements. Strain F24, which is pathogenic for melons, has four active Class II elements, *Fot1*, *Fot2*, *Hop* and *impala*, which have been identified by their insertion in the *niaD* gene (Daboussi and Langin 1994). One of them, *Fot1*, the first Class II transposon discovered in fungi (Daboussi et al. 1992), has recently been linked to the *pogo* superfamily (Robertson 1996; Smit and Riggs 1996). *Fot2* is a *Fot1*-like element, while *Hop* is a new type of transposon. The fourth, *impala*, is a member of the *Tc1-mariner* superfamily (Langin et al. 1995), which includes the *Tc1*-like family (TLEs) that was first identified in *Caenorhabditis elegans* (Emmons et al. 1983) and the *mariner*-like family (MLEs), initially described in *Drosophila mauritiana* (Jacobson et al. 1986). Members of this superfamily all have a TA insertion specificity and significant areas of sequence similarity, particularly in the region encoding the catalytic DDE domain (Doak et al. 1994). This domain is also present in the integrases of some retroviruses, where it was first identified, and in the transposases of bacterial elements (Fayet et al. 1990). One of them, the IS*630* bacterial IS family, contains a DDE region very similar to that of *Tc1-mariner* elements, suggesting that these elements have a common ancestor (Capy et al. 1996). This superfamily is widely distributed in protozoans (Jahn et al. 1993), and in animals (Avancini et al. 1996) as diverse as arthropods (Robertson 1993), platyhelminths (Garcia-Fernandez et al. 1995), fishes (Goodier and Davidson 1994), amphibians (Lam et al. 1996) and mammals (Auge-Gouillou et al. 1995). It appears to be absent from plants (P. Capy, unpublished results). *impala* was the first *Tc1-mariner* element identified in filamentous fungi, but two other fungal *Tc1-mariner* elements have since been discovered – *Ant-1* in *Aspergillus niger* (Glayzer et al. 1995) and *hupfer* in *Beauveria bassiana* (Maurer

et al. 1997) – indicating that this superfamily is also widespread in fungi. The *Tc1-mariner* superfamily includes a variety of related but distinct elements that fall into separate subfamilies, several of which may coexist in the same genome (Robertson and MacLeod 1993; Lam et al. 1996). The relationships between elements belonging to these subfamilies has been well studied in *mariner*-like elements (MLEs) (Robertson 1995).

The *impala*-160 copy (*imp160*) that transposed into the *niaD* gene, is 1280 nucleotides (nt) long, contains 27-bp inverted terminal repeats (ITRs) – the terminal nucleotides of which are similar to those of *Tc1* – and is inserted at a TA dinucleotide (Langin et al. 1995). Its central region contains three overlapping open reading frames (ORFs). A 340-amino acid (aa) polypeptide, similar in size to the transposases of the *Tc1-mariner* elements, can be derived from these ORFs but only after introducing two frameshifts; for this reason *imp160* was thought to be inactive. This putative polypeptide contains some blocks of amino acids that are similar to segments of *Tc1* or *mariner* transposases, depending on the regions compared. A phylogenetic analysis indicates that *impala* is closer to TLEs, albeit basally branched (Robertson and Asplund 1996). The *impala* element is represented by a small number of copies in strain F24. Variations both in the number of hybridizing bands in different restriction digests, and in signal intensity between the different copies suggest that this family is heterogeneous in structure. The existence in this strain of partially deleted copies or highly divergent copies could explain the variation of signal intensities. We have therefore cloned and sequenced the different copies in strain F24 to analyse the structure of this family and to identify autonomous elements. This analysis revealed that full-length and truncated copies coexist in this strain. Four copies, each encoding a putatively active transposase, are good candidates for autonomous elements. The copies can be grouped into three subfamilies, whose origins are discussed.

Materials and methods

Cloning of *impala* copies

Membranes replicated from a cosmid genomic DNA library previously prepared by partial *Sau*3 AI digestion (Diolez et al. 1993) were hybridized with an *imp160* probe. Positive clones were replicated and re-hybridized to confirm they were true positive clones. The different *impala* copies were subcloned by mixing 20 µl of each cosmid mini-preparation (Sambrook et al. 1989), digested to completion with *Eco*RI (Boerhinger), with 200 ng of plasmid pMLC28 (Levinson et al. 1984), digested with *Eco*RI and treated with alkaline phosphatase (Boehringer), following the manufacturers instructions. The DNA was then purified with phenol/chloroform, precipitated with ethanol and resuspended in 10 µl of 100 mM TRIS-HCl, 5 mM $MgCl_2$. It was ligated overnight at 16° C using the Amersham DNA Ligation system. DH5α bacteria were transformed and chloramphenicol-resistant colonies carrying *impala* inserts were identified by colony hybridization on a nylon filter (Amersham) using *impala* probe.

Sequencing and sequence analysis

DNA was sequenced manually by the dideoxy chain-termination method of Sanger et al. (1977), using a T7 and deaza G/A T7 sequencing kit (Pharmacia), or automatically, using a ABI Prism 373 sequencer and the ABIprism Dye Terminator DNA sequencing kit (Perkin Elmer Applied Biosystems). Automatic sequencing of *imp160* revealed that the two frameshifts interrupting the ORF at positions 329 and 690 (Langin et al. 1995) were in fact artifacts of GC compression and low resolution.

The *impala* elements were sequenced on both strands using a set of internal primers described by Langin et al. (1995) or with newly synthesized primers. The flanking regions were sequenced using primers located near, and directed towards, the end of the element.

ORFs were identified with DNA Strider 1.2. Sequences were aligned using ClustalV (Higgins and Sharp 1989) with final manual adjustment. A phylogenetic tree was built by the maximum parsimony method, using the Heuristic option of PAUP 3.0s (Swofford 1991). Homologs were searched for using the blastx option of the BLAST program. Secondary structure were predicted with PHDsec (Rost and Sanders 1993).

Southern blotting and hybridization

Fungal genomic DNA was extracted as described in Daboussi et al. (1989). Some 10 µg DNA was digested with *Eco*RI and fractionated in 0.7% agarose gels. Samples were blotted onto Nylon N-membrane (Amersham) and hybridized under standard conditions (Sambrook et al. 1989). Probes were labelled with ^{32}P by random priming using the Pharmacia T7 QuickPrime kit.

Polymerase Chain Reaction

PCR was done with 50 ng of plasmid in a 100-µl reaction volume, using Appligene buffer and 800 nM of each primer, 100 µM of each dNTP, 0.2 µl Appligene Taq polymerase, in a Braun thermocycler. Cycling conditions were: 1 min 30 at 94° C, 1 min at 60° C and 1 min 30 at 72° C for 30 cycles, followed by an elongation step of 15 min at 72° C. *imp160* was amplified using a single primer based on the ITRs (5′-CAGTGGGGTACAATAAGTTTG-3′). Flanking regions were amplified using universal or reverse primers (annealing to the vector) and customized internal primers.

Results

Cloning of F24 *impala* copies

Previous results revealed that the number of major hybridizing bands in a Southern blot of F24 genomic DNA varied with the enzyme used. *Eco*RI, an enzyme which does not cut within *imp160*, gave rise to six major bands that hybridized strongly with *imp160*, and at least two low-intensity bands. A cosmid library of strain F24 (Diolez et al. 1993) was screened using the entire *imp160* element as a probe to isolate the different copies. Out of 3840 clones 40 gave positive signals. DNA samples from 20 of them were digested with *Eco*RI and probed with *imp160*. The different cosmids were grouped into eight classes (1–8) depending on their restriction pattern (Table 1). Six classes (1, 3, and 5–8) contained cosmids having only one *Eco*RI fragment that hybridized to the probe, whereas classes 2 and 4 contained cosmids having two fragments that hybridized with different intensities. The ten *Eco*RI fragments were individually subcloned

Table 1 Cosmids used in this investigation

Cosmid classes	No. of cosmids per class	Size (in kb) of hybridizing *Eco*RI fragments[a]	Position of the inserts in the cosmids
1	5	5.5 (*impA*)	Internal
2	4	6.8 (*impB*), 8.0 (*impGΔ*)	Internal
3	2	8.4 (*impC*)	Internal
4	2	7.2 (*impD*), **9.0** (*impH*)	Internal
5	1	>12 (*impE*)	Internal
6	1	4.5 (*impF*)	Border
7	2	9.4 (*impG*)	Internal
8	3	**9.0** (*impH*)	Internal

[a] The names of the corresponding *impala* copies are indicated in *parentheses* after the size of the *Eco*RI fragments. The size of the duplicated fragments is shown in *bold*

into pMLC28. Six inserts were different in size and could be assigned to the six major *Eco*RI hybridizing bands in genomic DNA. These clones were named pImpE, pImpG, pImpC, pImpB, pImpD, and pImpA (Fig. 1 and Table 1). The large genomic fragment (about 20 kb) (Fig. 1), which gave a weak hybridization signal, did not correspond to any of the four remaining inserts on the basis of size. As three of them were internal to the cosmids, the only insert that could correspond to this large genomic fragment was the 4.5-kb insert, located at the border of cosmid 6 and probably truncated (Table 1). It was named pImpF and the truncation of this fragment was confirmed by sequencing. The three other fragments hybridized very weakly and could not be detected by Southern blotting. Two of them, in cosmids 4 and 8, were identical in size (9.0 kb) (Table 1). Digestion with *Bgl*II, which cuts five times within the two inserts, demonstrated that they had the same sequence, and they were named pImpH. These two fragments, both internal to the cosmids, were nevertheless the only fragments common to the two cosmids, suggesting that they correspond to the duplication of the same region. The last one, 8.0 kb long, was cloned from cosmid 2 and characterized by sequencing (see below).

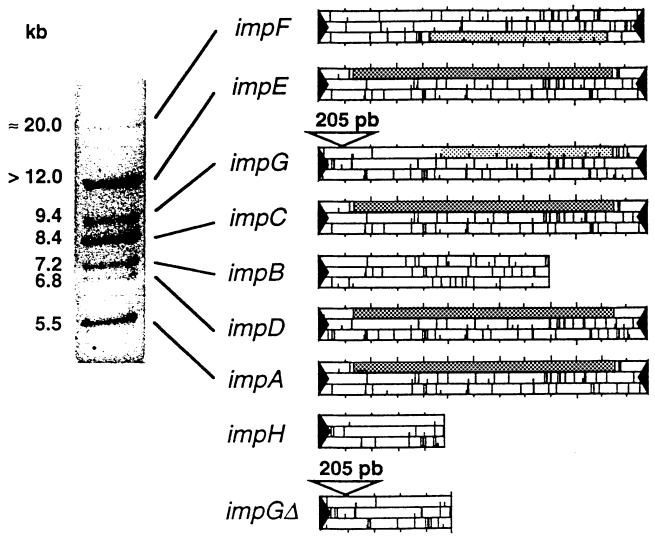


Fig. 1 Structures of *impala* copies. ITRs are represented by *black triangles*. ATG and stops are indicated in the three reading frames by *short* and *long vertical bars*, respectively. The major ORFs are in *dark grey*, the truncated ORF is in light grey. Insertion of the short sequence in *impG* is depicted *above* the element. The panel and the left shows a Southern blot of genomic DNA of strain F24 digested with *Eco*RI with assignment of the copies to the *imp160*-hybridizing fragments

Structure of *impala* copies

All the copies were sequenced (Fig. 1). Six copies – *impF*, *impE*, *impG*, *impC*, *impD* and *impA* – corresponded to complete elements with ITRs at both ends, delimiting a region of approximately 1280 nt, except of *impG* whose ITRs were 1500 nt apart. This larger size was due to the presence of a 210-bp insertion within the 5′ non-coding region. This short sequence had some characteristics of a solo-LTR (5-bp duplicated target site, short internal repeats, 5′-TG and CA-3′ terminal nucleotide motifs). *impB* and *impH* were both truncated in their 3′ regions, at positions 889 and 490, respectively. The 8.0-kb insert, subcloned from cosmid 2, contained a truncated *impala* copy corresponding to the first 500-bp in *impG*, and was called *impGΔ*. This analysis demonstrated that both full-length and 3′ truncated copies of *impala* can be found in *F. oxysporum*.

Analysis of the flanking regions of *impala* copies revealed that, like all members of the *Tc1-mariner* superfamily, *impala* copies are flanked by the dinucleotide TA (Fig. 2), except for *impD* and *impF*, which are bordered at the 5′ end by the dinucleotide CA. For the related *Tc3* element the TA has been demonstrated to result from the duplication of the target site (TSD) (van Luenen et al. 1994). The simplest explanation for the presence of a CA at the 5′ end of *impD* and *impF* is that a T was mutated to a C. As such a mutation in the TSD has never been reported for *Tc1-mariner* elements, it probably reflects ancient insertions rather than changes during transposition. No consensus for the flanking sequence of *impala* integration sites has been found, suggesting that the insertion specificity is restricted to TA.

Approximately 200–500 bp on each side of each of the 10 copies were sequenced. No large ORF was detected. Searching of sequence databases revealed no

TATATTATTTTCCTATCCCA**TA** *cagt.imp160.actg* **TA**CTCTCACAATGTCTACAACCG
TGCAGCTATTGTTAGCGTAG**TA** *cagt..impA..actg* **TA**TTGTAGCCAAAGAAGGTTTTA
CAAGGCCCCTTCCAACGTCA**TA** *cagt..impB*
ACAGGCAAACGTGCAGGAAA**TA** *cagt..impC..actg* **TA**GGCACTTACCTAATGTCGGAC
TTGAGTCCTAGAAAGTCCTA**CA** *cagt..impD..actg* **TA**CTACCAGAAAGCCTATGTATT
GACCGGTCGCACGCATTGCA**TA** *cagt..impE..actg* **TA**GGATCGTGATACAAGATAGAA
GTCAGTCAAATGCCCTAGAA**CA** *cagt..impF..actg* **TA**GGACAAAGAGCAACAACCTTA
CTTTAAATGAAACGGGTGAG**TA** *cagt..impG..actg* **TA**TCAGCGTTTACACAGCTAATA
ACTTGCCGCGATAATGCCAT**TA** *cagt..impH*

Fig. 2 5′ and 3′ flanking regions of the nine *impala* copies. *impala* sequences are in *italics*. The target site duplication is in *boldface*

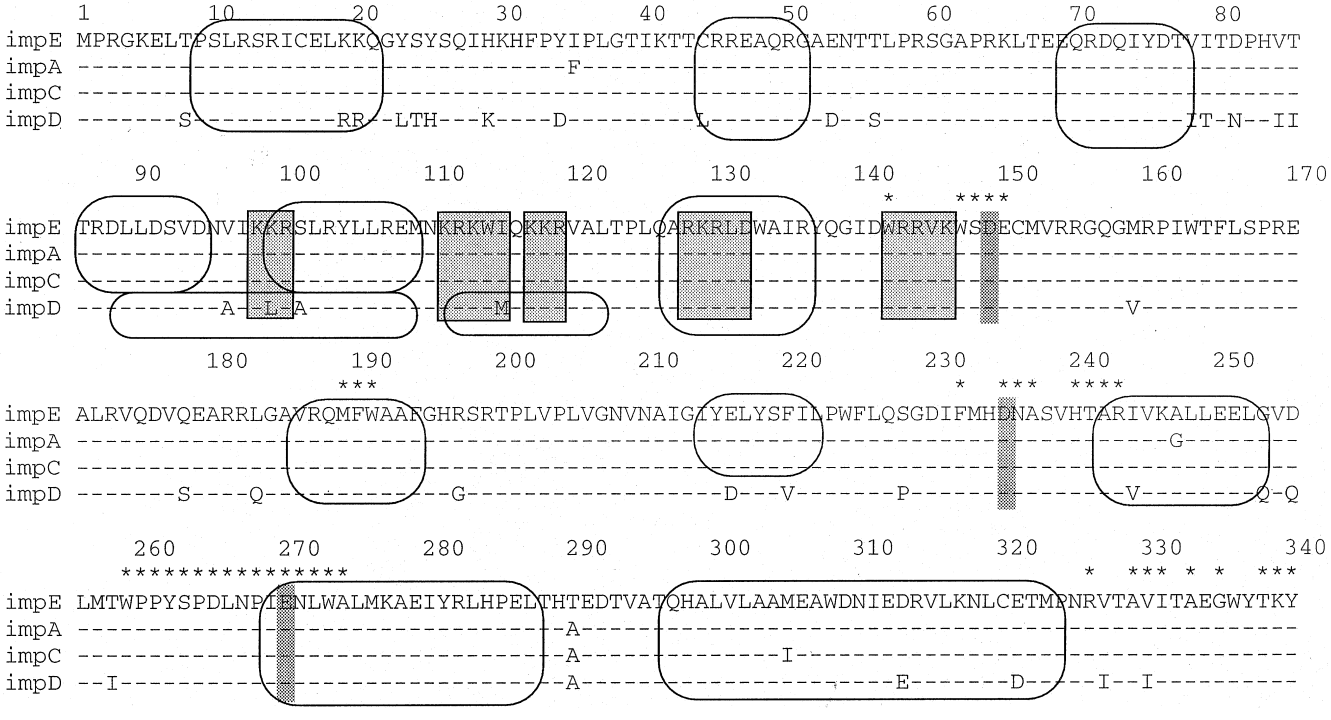
significant sequence similarity, except for the *impA* 5′ region. This element is inserted near an ORF that has significant similarity with several histidinol dehydrogenase genes. However, the insertion of *impA*, 116 bp downstream from the end of the coding sequence, probably has no effect on the putative activity of this gene. Several flanking regions of elements were amplified by PCR and used as probes on *F. oxysporum* genomic DNA, to examine the nature of the regions in which *impala* has inserted. This showed that *impA*, *impB*, *impC*, and *impE* are inserted in single-copy sequences, while *impF*, *impD*, *impG*, *and impH* are inserted in regions containing repeated sequences (data not shown).

Coding potential of *impala* copies

ORFs were searched for in all the elements. One uninterrupted ORF, extending from nt 133 to nt 1153 and potentially encoding a 340-amino acid polypeptide, was detected in four of the copies: *impA*, *impC*, *impD*, and *impE*. These putative transposases were similar in length to those described for elements belonging to the *Tc1-mariner* superfamily (Robertson 1995). An ORF potentially coding for the C-terminal part (233 aa) of the *impala* transposase, was also detected in *impF* and *impG*. No significant ORF was found in *impB* or *impH*, due to the presence of numerous stop codons. Alignment of the amino acid sequences revealed two groups of transposases (Fig. 3). The *impA*, *impC* and *impE* transposases were very similar, differing by only two or three amino acids. Nevertheless, there were more differences between the nucleotide sequences of *impA* and *impE* (14) than between those of *impC* and *impE* (2). The fourth copy, *impD*, appeared to be more divergent, with 36 amino acid changes, 16 being conservative changes relative to *impE*. Most of the nucleotide changes (94/154) detected in the ORF occurred in the wobble base of the codons.

Some functional domains are essential for transposase activity in *Tc1-mariner* elements. The DDE signature has two aspartate residues (D) and a glutamate residue (E). The second D is separated from the E by 34 or 35 amino acids. This domain was first identified in IS transposases and retroviral integrases (Fayet et al. 1990; Kulkosky et al. 1992), and has also been detected in the *Tc1-mariner* superfamily, although the MLEs have a third D instead of a E (Doak et al. 1994). This motif is thought to correspond to a cation binding site necessary for cleavage and strand transfer reactions. Mutation of any of these residues in retroviral integrase abolishes catalytic activity (Kulkosky et al. 1992). These are important for the activity of elements belonging to the *Tc1-mariner* superfamily, as shown for *Tc1* and *Tc3* (van Luenen et al. 1994; Vos and Plasterk 1994). Interestingly, the replacement of the last D by a E in the transposase of the *mariner* Mos1 autonomous element also leads to a loss of function (Lohe et al. 1997). The four *impala* putative transposases all have an intact DDE motif (Fig. 3).

Fig. 3 Amino acid sequences of the putative transposases encoded by four copies. Sequence identity with the *impE* sequence is indicated by the *dashes*. Changes are indicated by *letters*. *Asterisks* indicate amino acids that are conserved in *impala* and *Tc1* or *mariner*. The three conserved amino acids of the DDE domain are shown by vertical bars. *Grey boxes* correspond to the basic clusters of the putative bipartite NLS and *open boxes* indicate the α-helices



A nuclear localisation signal (NLS) conforming to the consensus bipartite NLS (two basic aminoacids followed by 10-aminoacid spacer and a cluster of five amino acids, of which three are basic (Dingwall and Laskey 1991), is present in the N-terminus of transposases encoded by TLEs, including *imp160*, and has been detected in the *mariner* Mos1 element (Ivics et al. 1996). This NLS is functional in fish TLEs (Ivics et al. 1996), and needed for the activity of the Mos1 copy of *mariner* (Lohe et al. 1997). In *impA*, *impC*, *impE*, and *imp160*, this motif is found at the same position, upstream of the DDE domain. An amino acid change has destroyed this motif in *impD* transposase, but other basic regions located downstream of this site might play the same role, although they fit the consensus defined by Dingwall and Laskey (1991) less well.

A third important domain, the DNA-binding domain, has been identified in N-terminal regions of some TLE transposases (Colloms et al. 1994; Vos and Plasterk 1994). This domain corresponds to a helix-turn-helix motif (HTH) identified in several *Tc1-mariner* transposase ORFs by multiple alignment analysis (Pietrokovski and Henikoff 1997). In *impala*, several conserved α-helices were detected in the N-terminal region. The first (residues 9–20) corresponds to helix 1 of *Tc3* transposase, which may be involved in the dimerization of the protein (van Pounderoyen et al. 1997); helices 4 and 5 (residues 86–94 and 99–111) correspond to the helix-turn-helix motif in block C of *mariner* transposase, that is similar to the HTH motif found in block A in TLEs (Pietrokovski and Henikoff 1997). This HTH motif binds to the ITRs, as determined from the crystal structure of the *Tc3* transposase (van Pouderoyen et al. 1997). In *impD*, helices 4 and 5 are fused, and an additional helix is found just dowstream. The HTH motif is thus present, but at a different position relative to the ORFs in other copies.

Thus, the four full-length copies encoding a putative transposase containing the intact functional domains required for transposition could all be autonomous.

Comparison of the sequences of *impala* elements defines three subfamilies

The DNA sequences were aligned for phylogenetic analysis. The *imp160* copy, corresponding to an additional copy specifically present in the *niaD160* mutant, was re-sequenced during this work and included in the alignment. The phylogenetic tree obtained by the maximum parsimony method showed that elements were clustered in three subfamilies (Fig. 4). The first, the E subfamily, contains *impA*, *impC*, *impE*, *impB*, and *imp160*, while *impD*, *impG* and *impH* comprise the second (D subfamily). The third, named the F subfamily, contains only one element, *impF*. Table 2 shows the pairwise divergences among the copies. The full-length copies in each subfamily were very similar (0.0 to 1.1%), whereas the two truncated copies, *impB* and *impH*, were more divergent (about 7 and 14% nucleotide polymorphism relative to their respective subfamilies). In subfamily E, *impE* and *imp160* are identical. Therefore, *imp160* which is the additional copy detected in the *niaD* gene in the mutant *niaD160* is probably the result of transposition of the *impE* copy. Nucleotide differences are equally distributed along the elements in this subfamily (Fig. 5B) and mainly represent transitions. This is particularly true for the truncated *impB* copy, in which transitions account for 97% of the nucleotide differences. These transitions are mainly G to A substitutions often located in YpG dinucleotides and, to a lesser extent, C to T transitions, all located in CpR dinucleotides. This bias was not observed in the D subfamily.

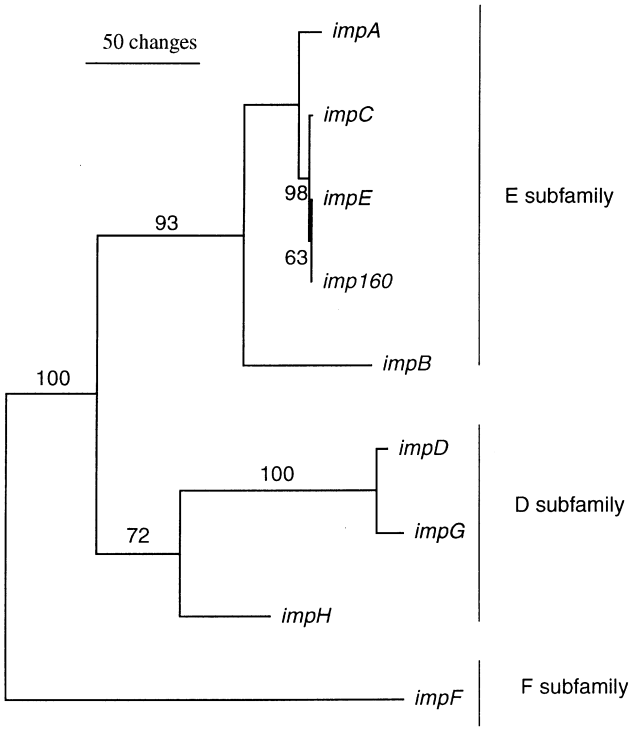


Fig. 4 Phylogenetic tree obtained by the maximum parsimony method. *impF* was used as the outgroup. The bootstrap values shown on the branches are expressed in percentages per 1000 repetitions

The subfamilies have quite different nucleotide sequences (17 to 29%) (see Table 2), but this does not preclude cross-hybridization. The two copies *impD* and *impF*, when used as probes, did not detect new copies in genomic DNA or new positive cosmids in the library. Sequence comparisons of representatives of the three subfamilies (*impE*, *impD* and *impF*) revealed that many of the differences are near the ends. For *impF,* the polymorphism extends over the first part of the coding region (Fig. 5A). But there are few nucleotide changes in the region containing the two functional domains (DDE and NLS), suggesting the existence of selective pressure on these domains, that could be indicative of recent activity of these elements.

We then examined these subfamilies to see if they showed specific structural characteristics. ITRs usually

Table 2 Pairwise divergence of the *impala* copies

Copy[a]	E subfamily					D subfamily			F subfamily
	imp160	*impE*	*impC*	*impA*	*impB*	*impD*	*impG*	*impH*	*impE*
imp160		0.00	0.15	1.09	6.64	17.49	17.80	23.01	24.12
impE	0		0.15	1.09	6.64	17.49	17.80	23.01	24.12
impC	2	2		1.09	6.64	17.49	17.80	23.01	24.12
impA	14	14	14		7.42	17.33	17.64	23.01	24.27
impB	59	59	59	66		23.28	23.73	28.31	29.13
impD	224	224	224	222	207		1.09	14.34	27.71
impG	228	228	228	226	211	14		14.86	28.10
impH	113	113	113	113	139	69	73		27.49
impE	309	309	309	311	259	355	360	135	

[a] Pairwise divergences are shown as percentages (*upper diagonal*) and number of nucleotide changes (*lower diagonal*). Divergences between elements were calculated for the whole element (about 1280 bp), except in the cases of *impB* (first 889 bp) and *impH* (first 490 bp)

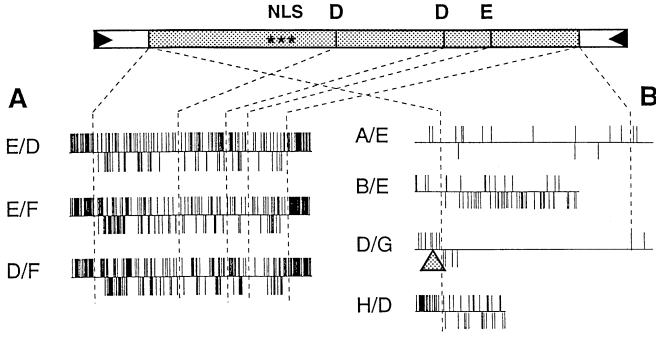


Fig. 5A, B Distribution of nucleotide polymorphism between (**A**) and within (**B**) subfamilies of *impala* elements. The structure of a typical copy is shown at the top. ITRs (*black triangles*), ORF (*grey box*), the three amino acids of the DDE signature and the putative NLS (*asterisks*) are indicated. Silent alterations within the ORF are shown as *below* the line vertical bars, and other changes are indicated *above* the line. The position of the 205-bp sequence inserted in *impG* is indicated by the *grey triangle below* the diagram

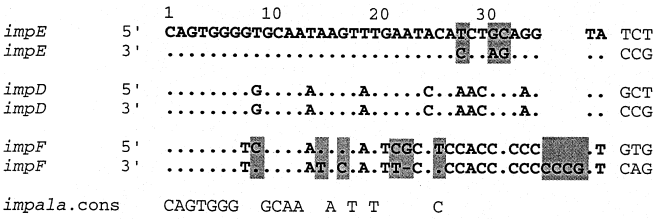


Fig. 6 Comparison of the ITRs. Inverted repeats are shown in *bold*. The *dots* indicate identity with the reference sequence (*impE*, left ITR). Differences are given by *letters*. Differences between left and right ITRs of each subfamily are *boxed in grey*. The *impala* ITR consensus is shown below

contain *cis* sequences important for transposition, especially the binding site for the transposase. Their divergence was therefore analysed (Fig. 6). *impD* contains perfect 37-bp ITRs and the ITRs of the E subfamily are perfect over a stretch of 27 bp. Allowing for three mismatches at positions 28, 31 and 32 leads to 37-bp imperfect repeats. The ITRs in *impF* are also 37 bp long and seem more divergent. The lone ITRs of the truncated elements appear to differ more from those of the members of their respective subfamilies. The ITRs from elements belonging to different subfamilies had quite divergent sequences, and these differences are not evenly distributed along the ITRs, since the first seven nucleotides are conserved. The first four are identical to those of *Tc1*, in which at least some of them are essential for in vitro strand cleavage (Vos and Plasterk 1994). This suggests these nucleotides are important for transposition. There are more differences in the sequences of the innermost part of the *impala* ITRs between subfamilies; so the sequence itself in this region seems to be less important than the ability of the two ITRs of an element to match. Mutational analysis of the inverted repeat of the bacterial transposon Tn*7* showed that, while some particular nucleotides of the ITRs are necessary for transposition, restoring inverted repeats in double mutants can lead to partial restoration of transposition (Tang et al. 1995).

Discussion

We have cloned and analysed the different copies of the *impala* transposable element detected in the genome of the *F. oxysporum* F24 strain. Ten full-length and truncated copies were isolated. The four truncated copies (all lacking their 3′ part), included two of the same element, lying in a duplicated region (*impH*). Another (*impGΔ*) resulted from the duplication of a part of a full-length copy (*impG*), which was itself interrupted in its 5′ part by a short sequence thought to be a solo-LTR. These elements were placed in three divergent subfamilies. The full-length members of each subfamily are more similar to each other than to the truncated copies. These truncated copies could be ancient insertions which have accumulated many mutations. For *impB*, belonging to the E subfamily, the high degree of sequence variation is due to many G to A and C to T transitions. Interestingly, its 3′ flanking region of 400 bp is extremely TA rich (82%). Such features are reminiscent of RIP (repeat-induced point mutation), a pre-meiotic inactivation process that occurs in the filamentous fungus *Neurospora crassa* (Cambareri et al. 1989). This process is very efficient and acts on duplicated sequences, at a precise time in the life

cycle (between fertilization and karyogamy). Inactivation by RIP is associated with methylation and with C to T and complementary G to A transitions, depending on the strand RIPped. RIP is restricted to duplicated sequences, but can extend to flanking regions (Foss et al. 1991). The bias in *impB* could perhaps result from RIP. However, several features do not fit RIP characteristics: first, *F. oxysporum* has no known sexual stage; second *impB* belongs to a subfamily containing several members, but is the only copy affected by this process; third, mutations in *impB* can be explained by C to T transitions occurring indiscriminately in CpG or CpA, while RIP occurs mainly in CpA; and lastly, the TA-rich region flanking *impB*, the sequence that could subjected to RIP, appears to be unique in the genome. However, intensive RIP-mediated mutagenesis leads to the rapid divergence of the duplicated sequences, which could prevent further cross-hybridization. In *Aspergillus fumigatus*, another asexual fungus, a repeated transposable element, *Afut1*, also appears to have suffered numerous CG to TA transitions, which, unlike the case in RIP, occur mainly in CpG dinucleotides (Neuveglise et al. 1996). As proposed for this latter element, the accumulation of CG to TA transitions in *impB* and its flanking region could result from a new phenomenon, somewhat different from RIP, or it could reflect the existence of something similar to RIP at a time when *F. oxysporum* still had a sexual stage.

Insertion specificity seems to be restricted to the TA target site duplication. No larger consensus sequence can be defined, unlike the case with the *C. elegans Tc1* element, which inserts preferentially in 12-bp consensus (Korswagen et al. 1996; Vos and Plasterk 1994) or for hymenopteran *mariner* elements which are found in large, degenerate palindromic sequences (Bigot et al. 1994). The *impala* elements are flanked by either single or repeated sequences. Curiously, the elements inserted in unique sequence contexts are all members of the E subfamily, whereas members of the D and F subfamilies are in repeated regions. Whether this is just coincidental or reflects a subfamily-specific insertion preference is unclear. Only a few insertion sites were analysed in this study. These insertions might reflect ancient selected transposition events, and the characteristics of *impala* insertion sites should become clearer after analysis of a large number of strains in which *impala* has recently transposed, as has been done for the *Tc1* and *Tc3* elements (van Luenen and Plasterk 1994).

Four full-length copies encode a putative 340-amino acid transposase that contains domains required for activity (NLS, DDE and DNA binding domain). Three belong to the E subfamily and differ by only a few amino acids. At least one of them (*impE*) is mobilisable, since it has transposed into the *niaD* gene, but we do not know whether it is transposed by its own transposase or if it is complemented in *trans* by another genomic copy. The fourth copy, *impD*, is very divergent from members of the E subfamily. However, many of the mutations that distinguish *impE* and *impD* are synonymous mutations, and numerous amino acid changes are conservative. This suggests that there have been selective constraints on evolution of the coding sequence, as might be expected for active genes. The recent activity of this copy is also suggested by analysis of its ITRs, which have identical 37-bp motifs. This copy may however be no longer mobilisable, because nucleotide changes in the TSD that replace the sequence TACAGT by CACAGT. Mutational analysis of the TSD of the related elements *Tc1* and *Tc3*, which have the same terminal sequence, have shown that a mutation of the TA and CG abolishes the endonucleotide cleavage of *Tc1* in vitro (Vos and Plasterk 1994). However, a mutation of TA to GC at both ends has no significant effect on *Tc3* transposition (van Luenen et al. 1994). Two other copies contain an ORF that could encode a protein lacking the first 108 residues. The DNA-binding domain is probably located in this part of the transposase, as for *Tc1* and *Tc3* (Colloms et al. 1994; Vos and Plasterk 1994), so the truncated *impala* transposases might be defective for transposition. The sequence surrounding the putative initiation codon ATG is compatible with the consensus defined for filamentous fungi (Gurr et al. 1987). These truncated proteins could affect the transposition frequency, as has been shown for a naturally deleted *P* element that encodes a truncated KP protein that acts as a negative regulator or *P* element transposition (Black et al. 1987).

A biological test is required to demonstrate the activity of *impala* copies. Activity or inactivity of different *mariner* copies has previously been demonstrated in *D. melanogaster* by transcomplementation of an inactive element, *peach*, inserted in the *white* gene (Maruyama et al. 1991). A similar test, based on the nitrate reductase gene, has been developed in our laboratory and used to demonstrate the autonomy of copies of *Fot1*, another *F. oxysporum* element (Migheli et al., submitted). The phenotypic assay will allow us to identify active and inactive *impala* copies. Transposases carried by elements from the E subfamily contain few variations in amino acids. Some of them may be functionally important and this could be deduced by comparing the activities of the different transposases, as for *mariner* elements (Maruyama et al. 1991; Lohe et al. 1997). This test could also be used to determine whether two very divergent copies such as *impE* and *impD*, can transcomplement, as seen for *P* elements differing by 24% (Simonelig and Anxolabéhère 1991).

The presence of three distinct subfamilies raises the question of the dynamics and evolution of the *impala* family. Several subfamilies have also been described for MLEs and TLEs (Robertson and Asplund 1996). For MLEs, at least five major subfamilies, which can coexist in the same species (i.e. in *Ceratitis capitata*), have been identified (Robertson and MacLeod 1993). The *impala* elements described in this study are present in the same strain, whereas MLE analyses were done on elements sometimes belonging to very distantly related species. These two families represent then two different situations.

However, several parallels can be drawn from the comparison of the two families. For both, the subfamilies are defined by the distinct clades they form on phylogenetic trees. The divergence within each *mariner* subfamily, which can be relatively high when copies from distantly related species are compared, does not exceed 10% (Robertson 1993) – generally 1–2% – in *Drosophila* species (Capy et al. 1992; Maruyama and Hartl 1991b; Medhora et al. 1991) – for elements in the same species, as seen for *impala* (about 1%, but reaching 14% for the truncated copies). Between subfamilies, nucleotide divergence is higher for MLEs – typically 45 to 60% – (Robertson 1993), than for *impala* (24% on average).

The widespread distribution of MLEs within insects suggests that this family is an ancient component of their genomes, that was present before the diversification of insects (Capy et al. 1994). Its evolution is believed to be the consequence of three types of mechanisms: vertical inactivation, stochastic loss and occasional horizontal transmission (Lohe et al. 1995). *impala* could conform to the same evolutionary scheme. Inactive *impala* copies have been identified. This family seems to be present in most other strains of *F. oxysporum* (unpublished data), but nothing is known about its presence in other species. The simplest hypothesis is that *impala* is an ancient element that has undergone two amplifications, an ancient one that gave rise to the three present subfamilies and a more recent one that led to the different copies within E and D subfamilies. There could also have been successive introductions of *impala* elements by horizontal transfer. Horizontal transfer has been invoked for several transposable elements: *P* from *Drosophila melanogaster* (Kidwell 1992), but also for MLEs for which the molecular phylogenies are sometimes not congruent with their host phylogenies (Maruyama and Hartl 1991a; Garcia-Fernandez et al. 1995; Robertson and Lampe 1995). Examination of the distribution of *impala* and its subfamilies in other *F. oxysporum* strains will help to define the types of mechanisms responsible for the present situation. The presence of the different subfamilies in several *formae speciales* would favor their ancient origin and vertical transmission of the subfamilies, while the presence of one subfamily restricted to a small number of strains, closely related to F24, would favor horizontal transfer.

The results indicate that the active *impala* family is composed of full-length and truncated copies. The insertion specificity is low. The activity of the four putatively autonomous copies might be assessed in a biological test similar to that developed for *Fot1*. Previous results have shown that excision of *impala* is frequently associated with its reinsertion at another genomic position. Hence *impala* might be a powerful tool for tagging genes, at least in *F. oxysporum*, a species in which the absence of sexuality makes gene isolation very difficult.

Acknowledgements We thank Fiano Kaper and Donal O'Sullivan for critical reading of the manuscript, Catherine Gerlinger for technical assistance, and Owen Parkes for editing the English text.

References

Auge-Gouillou C, Bigot Y, Pollet N, Hamelin MH, Meunier RM, Periquet G (1995) Human and other mammalian genomes contain transposons of the *mariner* family. FEBS Lett 368:541–546

Avancini RM, Walden KK, Robertson HM (1996) The genomes of most animals have multiple members of the *Tc1* family of transposable elements. Genetica 98:131–140

Bigot Y, Hamelin MH, Capy P, Periquet G (1994) *Mariner*-like elements in hymenopteran species: insertion site and distribution. Proc Natl Acad Sci USA 91:3408–3412

Black DM, Jackson MS, Kidwell MG, Dover GA (1987) *KP* elements repress *P*-induced hybrid dysgenesis in *Drosophila melanogaster*. EMBO J 6:4125–4135

Cambareri EB, Jensen BC, Schabtach E, Selker EU (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. Science 244:1571–1575

Capy P, Koga A, David JR, Hartl DL (1992) Sequence analysis of active *mariner* elements in natural populations of *Drosophila simulans*. Genetics 130:499–506

Capy P, Langin T, Bigot Y, Brunet F, Daboussi MJ, Periquet G, David JR, Hartl DL (1994) Horizontal transmission versus ancient origin: *mariner* in the witness box. Genetica 93:161–170

Capy P, Vitalis R, Langin T, Higuet D, Bazin C (1996) Relationships between transposable elements based upon the integrase-transpose domains: is there a common ancestor? J Mol Evol 42:359–368

Colloms SD, van LH, Plasterk RH (1994) DNA binding activities of the *Caenorhabditis elegans Tc3* transposase. Nucleic Acids Res 22:5548–5554

Daboussi MJ, Langin T (1994) Transposable elements in the fungal plant pathogen *Fusarium oxysporum*. Genetica 93:49–59

Daboussi MJ, Djeballi A, Gerlinger C, Blaiseau PL, Bouvier I, Cassan M, Lebrun MH, Parisot D, Brygoo Y (1989) Transformation of seven species of filamentous fungi using the nitrate reductase gene of *Aspergilus nidulans*. Curr Genet 15:453–456

Daboussi MJ, Langin T, Brygoo Y (1992) *Fot1*, a new family of fungal transposable elements. Mol Gen Genet 232:12–16

Dingwall C, Laskey RA (1991) Nuclear targeting sequences–a consensus? Trends Biochem Sci 16:478–481

Diolez A, Langin T, Gerlinger C, Brygoo Y, Daboussi MJ (1993) The *nia* gene of *Fusarium oxysporum*: isolation, sequence and development of a homologous transformation system. Gene 131:61–67

Doak TG, Doerder FP, Jahn CL, Herrick G (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci USA 91:942–946

Emmons SW, Yesner L, Ruan KS, Katzenberg D (1983) Evidence for a transposon in *Caenorhabditis elegans*. Cell 32:55–65

Fayet O, Ramond P, Polard P, Prere MF, Chandler M (1990) Functional similarities between retroviruses and the *IS3* family of bacterial insertion sequences? Mol Microbiol 4:1771–1777

Foss EJ, Garrett PW, Kinsey JA, Selker EU (1991) Specificity of repeat-induced point mutation (RIP) in *Neurospora*: sensitivity of non-*Neurospora* sequences, a natural diverged tandem duplication, and unique DNA adjacent to a duplicated region. Genetics 127:711–717

Garcia-Fernandez J, Bayascas RJ, Marfany G, Munoz MA, Casali A, Baguna J, Salo E (1995) High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthes): evidence for a trans-phyla horizontal transfer. Mol Biol Evol 12:421–431

Glayzer DC, Roberts IN, Archer DB, Oliver RP (1995) The isolation of *Ant1*, a transposable element from *Aspergillus niger*. Mol Gen Genet 249:432–438

Goodier JL, Davidson WS (1994) *Tc1* transposon-like sequences are widely distributed in salmonids. J Mol Biol 241:26–34

Gurr SJ, Unkles SE, Kinghorn JR (1987) The structure and organizaton of nuclear genes for filamentous fungi. In: King-

horn JR (ed) Gene structure in eukaryotic microbes. Academic Press, Oxford Washington DC, pp 93–139
Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. Comput Appli Biosci 5:151–153
Ivics Z, Izsvak Z, Minter A, Hackett PB (1996) Identification of functional domains and evolution of *Tc1*-like transposable elements. Proc Natl Acad Sci USA 93:5008–5013
Jacobson JW, Medhora MM, Hartl DL (1986) Molecular structure of a somatically unstable transposable element in *Drosophila*. Proc Natl Acad Sci USA 83:8684–8688
Jahn CL, Doktor SZ, Frels JS, Jaraczewski JW, Krikau MF (1993) Structures of the *Euplotes crassus Tec1* and *Tec2* elements: identification of putative transposase coding regions. Gene 133:71–78
Kidwell MG (1992) Horizontal transfer of *P* elements and other short inverted repeat transposons. Genetica 86:275–286
Korswagen HC, Durbin RM, Smits MT, Plasterk RH (1996) Transposon *Tc1*-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. Proc Natl Acad Sci USA 93:14680–14685
Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol 12:2331–2338
Lam WL, Seo P, Robison K, Virk S, Gilbert W (1996) Discovery of amphibian *Tc1*-like transposon families. 257:359–366
Langin T, Capy P, Daboussi MJ (1995) The transposable element *impala*, a fungal member of the *Tc1-mariner* superfamily. Mol Gen Genet 246:19–28
Levinson A, Silver D, Seed B (1984) Minimal size plasmids containing an M13 origin for production of single-strand transducing particles. J Mol Appl Genet 2:507–517
Lohe AR, De Aguiar D, Hartl DL (1997) Mutations in the *mariner* transposase: the D, D(35)E consensus sequence is nonfunctional. Proc Natl Acad Sci USA 94:1293–1297
Lohe AR, Moriyama EN, Lidholm DA, Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. Mol Biol Evol 12:62–72
Maruyama K, Hartl DL (1991a) Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. J Mol Evol 33:514–524
Maruyama K, Hartl DL (1991b) Evolution of the transposable element *mariner* in *Drosophila* species. Genetics 128:319–29
Maruyama K, Schoor KD, Hartl DL (1991) Identification of nucleotide substitutions necessary for trans-activation of *mariner* transposable elements in *Drosophila*: analysis of naturally occurring elements. Genetics 128:777–784
Maurer P, Réjasse A, Capy P, Langin T, Riba G (1997) Isolation of the transposable element *hupfer* from the entomopathogenic fungus *Beauveria bassiana* by insertion mutagenesis in the nitrate reductase structural gene. Mol Gen Genet 256:195–202
Medhora M, Maruyama K, Hartl DL (1991) Molecular and functional analysis of the *mariner* mutator element Mos 1 in *Drosophila*. Genetics 128:311–318
Neuveglise C, Sarfati J, Latge JP, Paris S (1996) *Afutl*, a retrotransposon-like element from *Aspergillus fumigatus*. Nucleic Acids Res 24:1428–1434
Pietrokovski S, Henikoff S (1997) A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. Mol Gen Genet 254:689–695
Robertson HM (1993) The *mariner* transposable element is widespread in insects. Nature 362:241–245
Robertson HM (1995) The *mariner-Tc1* superfamily of transposons in animals. J Insect Physiol 41:99–105
Robertson HM (1996) Members of the *pogo* superfamily of DNA-mediated transposons in the human genome. Mol Gen Genet 252:761–766
Robertson HM, Asplund ML (1996) *Bmmar1*: a basal lineage of the *mariner* family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem Mol Biol 26:945–954
Robertson HM, Lampe DJ (1995) Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Mol Biol Evol 12:850–862
Robertson HM, MacLeod EG (1993) Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol Biol 2:125–139
Rost B, Sanders C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 232:584–599
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Simonelig M, Anxolabehere D (1991) A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. Proc Natl Acad Sci USA 88:6102–6106
Smit AF, Riggs AD (1996) *Tiggers* and DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 93:1443–1448
Swofford DL (1991) PAUP: phylogenetic analysis using parsimony, version 3.0s. Illinois Natural History Survey, Champaign, Ill
Tang Y, Cotterill S, Lichtenstein CP (1995) Genetic analysis of the terminal 8-bp inverted repeats of transposon Tn*7*. Gene 162:41–46
van Luenen H, Colloms SD, Plasterk RH (1994) The mechanism of transposition of *Tc3* in *C. elegans*. Cell 79:293–301
van Luenen H, Plasterk RH (1994) Target site choice of the related transposable elements *Tc1* and *Tc3* of *Caenorhabditis elegans*. Nucleic Acids Res 22:262–269
van Pouderoyen G, Ketting RF, Perrakis A, Plasterk RH, Sixma TK (1997) Crystal structure of the specific DNA-binding domain of *Tc3* transposase of *C. elegans* in complex with transposon DNA. EMBO J 16:6044–6054
Vos JC, Plasterk RH (1994) *Tc1* transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. EMBO J 13:6125–6132