Microbiology (2013), **159**, 112–125 **DOI** 10.1099/mic.0.062901-0

Boto, a class II transposon in *Moniliophthora perniciosa*, is the first representative of the *PIF/Harbinger* superfamily in a phytopathogenic fungus

Jorge Fernando Pereira,[1]† Ana Paula Morais Martins Almeida,[1]‡ Júnio Cota,[1]§ João Alencar Pamphile,[2] Gilvan Ferreira da Silva,[1]‖ Elza Fernandes de Araújo,[1] Karina Peres Gramacho,[3] Sérgio Hermínio Brommonschenkel,[4] Gonçalo Amarante Guimarães Pereira[5] and Marisa Vieira de Queiroz[1]

Correspondence
Marisa V. de Queiroz
mvqueiro@ufv.br

[1]Universidade Federal de Viçosa, Departamento de Microbiologia, CEP 36571-000, Viçosa, MG, Brazil

[2]Universidade Estadual de Maringá, Departamento de Biologia Celular e Genética, CEP 87020-900, Maringá, PR, Brazil

[3]CEPLAC, Centro de Pesquisa do Cacau, CEP 45600-000, Itabuna, BA, Brazil

[4]Universidade Federal de Viçosa, Departamento de Fitopatologia, CEP 36571-000, Viçosa, MG, Brazil

[5]Universidade Estadual de Campinas, Departamento de Genética e Evolução, CEP 13083-970, Campinas, SP, Brazil

Boto, a class II transposable element, was characterized in the *Moniliophthora perniciosa* genome. The *Boto* transposase is highly similar to plant *PIF*-like transposases that belong to the newest class II superfamily known as *PIF*/*Harbinger*. Although *Boto* shares characteristics with *PIF*-like elements, other characteristics, such as the transposase intron position, the position and direction of the second ORF, and the footprint, indicate that *Boto* belongs to a novel family of the *PIF*/*Harbinger* superfamily. Southern blot analyses detected 6–12 copies of *Boto* in C-biotype isolates and a ubiquitous presence among the C- and S-biotypes, as well as a separation in the C-biotype isolates from Bahia State in Brazil in at least two genotypic groups, and a new insertion in the genome of a C-biotype isolate maintained in the laboratory for 6 years. In addition to PCR amplification from a specific insertion site, changes in the *Boto* hybridization profile after the *M. perniciosa* sexual cycle and detection of *Boto* transcripts gave further evidence of *Boto* activity. As an active family in the genome of *M. perniciosa*, *Boto* elements may contribute to genetic variability in this homothallic fungus. This is the first report of a *PIF*/*Harbinger* transposon in the genome of a phytopathogenic fungus.

Received 15 August 2012
Revised 22 October 2012
Accepted 24 October 2012

†Present address: Embrapa Trigo, Rodovia BR 285 Km 294, CEP 99001-970, Passo Fundo, RS, Brazil.

‡Present address: Universidade Federal de Minas Gerais, Departamento de Microbiologia, Av. Antônio Carlos 6627, CEP 31270-901, Belo Horizonte, MG, Brazil.

§Present address: Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Rua Giuseppe Máximo Scolfaro 10000, CEP 13083-970, Campinas, SP, Brazil.

‖Present address: Embrapa Amazônia Ocidental, Rodovia AM-010 Km 29, CEP 69010-970, Manaus, AM, Brazil.

Abbreviations: IS, insertion sequence; MITE, miniature inverted-repeated transposable element; TIR, terminal inverted repeat; TSD, target site duplication.

The GenBank/EMBL/DDBJ accession number for the *Boto* sequence reported in this paper is EU218539.

INTRODUCTION

Eukaryotic transposable elements are divided into two main categories according to their transposition mechanism: the class I elements that transpose by an intermediate RNA and are further divided into the five orders LTR, DIRS, Penelope-like, LINEs and SINEs (Wicker *et al.*, 2007); and the class II elements that transpose directly at the DNA level, not requiring an RNA transposition intermediate. Class II elements can be further divided into subclasses, superfamilies and families by the transposition mechanisms and structural features of the terminal inverted repeats (TIRs), the transposase and the target site duplication (TSD) (Daboussi & Capy, 2003; Wicker *et al.*, 2007). Class II elements belonging to the

superfamilies *Tc1/mariner*, *hAT*, *mutator* and MITEs (miniature inverted-repeated transposable elements) have already been identified in several species of filamentous fungi (Daboussi & Capy, 2003). Moreover, as new elements are described and new eukaryotic genomes are sequenced, new groups of elements are identified (Goodwin & Poulter, 2001; Goodwin *et al.*, 2003).

One of the 10 class II superfamilies identified so far in eukaryotic organisms is the *PIF/Harbinger* superfamily. The first two elements described in this superfamily were the *PIF* element (P instability factor) of maize (Walker *et al.*, 1997) and the *Harbinger* element of *Arabidopsis thaliana* (Kapitonov & Jurka, 1999). *PIF/Harbinger* elements share characteristics with other groups of transposons, such as the small TIRs and the 3 bp TSD. However, some unique characteristics distinguish *PIF/Harbinger* elements from other superfamilies: (i) the presence of two open reading frames (ORFs), one coding for a transposase and the other for a protein of unknown function but showing weak similarity to *myb* transcription factors (Jiang *et al.*, 2003); (ii) a distant relationship between the *PIF/Harbinger* transposase and the transposase of bacterial insertion sequences (IS) of the IS*5* group; and (iii) their direct link in origin and mobility of non-autonomous MITEs (Zhang *et al.*, 2001, 2004; Grzebelus *et al.*, 2006). The *Harbinger* and *PIF* elements, in addition to the rice element named *Pong* (Zhang *et al.*, 2004), can be seen as the founding members of this widespread superfamily of DNA transposons. A distribution analysis identified more than 600 *PIF*-like transposases from 35 species of plants and 19 species of animals (Zhang *et al.*, 2004), and different *PIF/Harbinger* families have been found in protists, plants, insects, worms and vertebrates (Jurka & Kapitonov, 2001; Kapitonov & Jurka, 2004; Grzebelus *et al.*, 2006; Zhou *et al.*, 2010, 2012). Curiously, sequences similar to *PIF*-like elements were reported in only two species of fungi, *Cryptococcus neoformans* and *Neurospora crassa* (Zhang *et al.*, 2001, 2004). This observation is interesting because a great number of transposable elements from varying superfamilies have been identified in fungal genomes (Wöstemeyer & Kreibich, 2002; Daboussi & Capy, 2003; Pereira *et al.*, 2006).

Mutagenic effects of transposons could be one of the main mechanisms responsible for the high adaptability and plasticity exhibited by numerous species of pathogenic fungi (Daboussi & Capy, 2003; Shnyreva, 2003; Pereira *et al.*, 2006; Schmidt & Panstruga, 2011). In this context, studying transposable elements in the plant pathogen *Moniliophthora* (formerly *Crinipellis*) *perniciosa*, the causal agent of witches' broom disease of cacao, is important to understand the mechanisms related to genetic variability in this species. This fungus attacks cacao plantations in South and Central America and represents the main threat in south-eastern Bahia, the main cacao-producing region in Brazil (Pereira *et al.*, 1996). In addition to cacao (*Theobroma cacao*), *M. perniciosa* has other plant hosts, and a classification based on pathological data divides the species into the following three biotypes: the C-biotype infects species of the family Sterculiaceae (Evans, 1978; Bastos *et al.*, 1988), the S-biotype infects plants of the family Solanaceae (Bastos & Evans, 1985; Bastos *et al.*, 1988) and the L-biotype is a saprotroph that colonizes a wide variety of substrates (Evans, 1978; Hedger *et al.*, 1987). The genetic variability of *M. perniciosa* has been evaluated through different molecular studies that revealed a high degree of variability among isolates of this species (Andebrhan & Furtek, 1994; Andebrhan *et al.*, 1999; de Arruda *et al.*, 2003a, b; Rincones *et al.*, 2003, 2006; Ploetz *et al.*, 2005).

In the present work, we describe the isolation and characterization of a class II transposable element in the *M. perniciosa* genome. This element, called *Boto*, is the first representative of the *PIF/Harbinger* superfamily identified in a phytopathogenic fungus.

METHODS

Fungal strains and growth conditions. Isolates of *M. perniciosa* examined in the present study are listed in Table 1. Basidiomata from isolate 1919 were obtained from mycelial mats as described by Griffith & Hedger (1993) with the modifications introduced by Niella *et al.* (1999).

Isolation of recombinant phages. A sequence showing similarity to plant *PIF*-like transposase (e-value 1×10^{-29}) was obtained from the database of the Witches' Broom Genome Project. Primers CPORT1 (5′-TTGCTTGTGAGCTTGGTGTC) and CPORT2 (5′-GCCTGAGCATGTCGAAGATT) were used to amplify a 795 bp fragment corresponding to part of the transposase coding region that was subsequently used as a probe for the isolation of recombinant phages from a genomic library of *M. perniciosa* cloned into the λEMBL3 bacteriophage (Benton & Davis, 1977). Hybridizations were conducted at 65 °C using the Gene Images Random Primer Labelling Module and the CDP-Star Detection Module (Amersham Biosciences) according to the manufacturer's instructions. The plates containing the positive phages were individually collected, and second and third screenings were conducted using the same conditions described above. DNA was extracted from the positive phages following the protocol described by Felipe *et al.* (1992). Cleavage of the phage DNA was performed using different restriction enzymes according to the manufacturer's instructions.

Cloning of the *Boto* element, sequencing and sequence analysis. Fragments generated from the digestion of the isolated phages were cloned into the pBluescript II KS+ vector (Stratagene). DNA sequencing was performed according to the dideoxynucleotide chain-termination method (Sanger *et al.*, 1977) in a MegaBACE 500 sequencer (Amersham Biosciences). Analyses of DNA and protein sequences were performed using the BLAST algorithm (Altschul *et al.*, 1997), CLUSTAL W program (Thompson *et al.*, 1994), the CD-Search program to identify conserved domains (Marchler-Bauer & Bryant, 2004) and the AUGUSTUS program for gene prediction (Stanke & Morgenstern, 2005).

Phylogenetic analysis. The sequences of the fungal, oomycete, plant and animal *PIF/Harbinger* transposase proteins were obtained from GenBank. The sequences were aligned using the CLUSTAL W program, and phylogenetic analyses were performed based on the

Table 1. Isolates of *Moniliophtora perniciosa* used in this study

Isolate number	Isolate identification	Biotype	Chromosomal group*	Location†	Host
1	FA551	C	–	Tabatinga/AM	*Theobroma* sp.
2	ESJOH1	C	–	Marituba/PA	*Theobroma cacao*
3	ESJOH2	C	–	Ouro Preto do Oeste/RO	*Theobroma cacao*
4	ESJOH3	C	–	Belém/PA	*Theobroma cacao*
5	CP02-1	C	CP-C1	Itajaípe/BA	*Theobroma cacao*
6	Belmont	C	CP-C1	Belmonte/BA	*Theobroma* sp.
7	Ilhéus	C	CP-C1	Ilhéus/BA	*Theobroma* sp.
8	FA563	C	CP-C1	Itabuna/BA	*Theobroma cacao*
9	Santo Amaro	C	CP-C2	Santo Amaro/BA	*Theobroma* sp.
10	FA42	C	CP-C2	Itabuna/BA	*Theobroma cacao*
11	FA276	C	CP-C2	Itabuna/BA	*Theobroma cacao*
12	FA293	C	–	Gandu/BA	*Theobroma cacao*
13	FA562	C	CP-C2	Itabuna/BA	*Theobroma cacao*
14	Lep1	L	–	Pichilingue/Ecuador	*Arrabidaea verrucosa*
15	SCFT	L	–	San Carlos/Ecuador	*Arrabidaea verrucosa*
16	SCL4	L	–	San Carlos/Ecuador	*Arrabidaea verrucosa*
17	FA607	S	–	Coimbra/MG	*Solanum lycocarpum*
18	FA609	S	–	Poços de Caldas/MG	*Solanum* sp.
19	DOA-105	S	–	Jataí/GO	*Solanum lycocarpum*
20	LA17	L	–	Pichilingue/Ecuador	*Arrabidaea verrucosa*
21	RWB500	S	–	Mariana/MG	*Solanum cernum*
22	RWB551	S	–	Juiz de Fora/MG	*Solanum lycocarpum*
23	FA277	C	–	Itabuna/BA	*Theobroma cacao*
24	FA281	C	CP-C2	Aiquara/BA	*Theobroma cacao*
25	DOA100	C	–	–	*Theobroma cacao*
26	CP02‡	C	CP-C1	Itajaípe/BA	*Theobroma cacao*
27	ALF42	C	–	Itabuna/BA	*Theobroma cacao*
28	ALF110	C	–	–	*Theobroma cacao*
29	ALF276	C	–	Itabuna/BA	*Theobroma cacao*
30	ALF277	C	–	Itabuna/BA	*Theobroma cacao*
31	ALF278	C	–	Itabuna/BA	*Theobroma cacao*
32	ALF301	C	–	–	*Theobroma cacao*
33	ALF305	C	–	–	*Theobroma cacao*
34	ALF321	C	–	Ilhés/BA	*Theobroma cacao*
35	606GD-W	C	–	Itabuna/BA	*Theobroma cacao*
36	676GD-W	C	–	Floresta Azul/BA	*Theobroma cacao*
37	896FD-W	C	–	Jaguaquara/BA	*Theobroma cacao*
38	948FD-W	C	–	–	*Theobroma cacao*
39	1734D-W	C	–	Gandu/BA	*Theobroma cacao*
40	FA317	C	–	–	*Theobroma cacao*
41	SABA	C	–	–	*Theobroma cacao*
–	1919	C	–	–	*Theobroma cacao*

*Chromosomal groups 1 or 2 determined according to Rincones *et al.* (2006).
†AM, Amazonas; BA, Bahia; GO, Goiás; MG, Minas Gerais; PA, Pará; RO, Rondônia.
‡Isolate CP02 was used in the Witches' Broom Genome Project.

neighbour-joining method (Saitou & Nei, 1987) using bootstrap values based on 1000 replicates.

Footprint analysis. Primers Boto2.1 (5′-TGTAGGCATTCGGACTTTCGG) and Boto2.2 (5′-TTCGGATGCTCTTGCCGT) were designed based on the *Boto* flanking regions present in the λ phage 2.1.1. The expected 185 bp PCR fragment was precipitated and used for sequencing as described above.

DNA extraction and PCR amplification. Total DNA was extracted as described by Specht *et al.* (1982). The PCR amplification was performed in a thermocycler (PTC-100; MJ Research) with the following programme: for primers CPORT1 and CPORT2, 40 cycles of 1 min at 94 °C, 1 min at 55 °C and 1 min at 72 °C, and a final extension step at 72 °C for 10 min; and for primers Boto2.1 and Boto2.2, 30 cycles of 30 s at 94 °C, 30 s at 62 °C and 30 s at 72 °C, and a final extension step at 72 °C for 3 min. The reactions were carried out

in a final volume of 25 μl containing 1× thermophilic DNA poly Buffer (Promega), 2.5 mM $MgCl_2$, 400 μM dNTPs, 0.2 μM each primer, 50 ng total DNA and 1 unit *Taq* DNA Polymerase (Promega).

Southern hybridization analysis. For phage characterization, the viral DNA (2 μg) was digested with the restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I and *Sal*I (data not shown). After the sexual cycle in the *M. perniciosa* isolates, the distribution, copy number analyses and hybridization profiles were performed with total DNA (3 μg) digested with *Hin*dIII or *Sal*I. These enzymes do not cut inside the 795 bp transposase fragment used as the probe. The digested DNA was then electrophoresed in a 0.7 % agarose gel and transferred to a Duralon-UV nylon membrane (Stratagene) following standard procedures (Sambrook *et al.*, 1989). *Hin*dIII-cleaved DNA was hybridized at 58 °C but that temperature exhibited low specificity for the *Sal*I-cleaved DNA, making the results difficult to interpret. Subsequently, hybridization with the *Sal*I-cleaved DNA was performed at 65 °C. Probe labelling, hybridization and detection were performed with the Gene Images Random Primer Labelling Module and the CDP-Star Detection Module (Amersham Biosciences) according to the manufacturer's instructions.

RNA extraction, RT-PCR analysis and cDNA cloning. To obtain the *M. perniciosa* mycelial mass for the RT-PCR experiment, five mycelial discs (7 mm each) were placed in PDA medium at 27 °C for 10 days. Once grown, 10 mycelial discs were cut into smaller fragments and transferred to 125 ml Erlenmeyer flasks containing 50 ml Pontecorvo's minimal medium (Pontecorvo *et al.*, 1953) and incubated for 7 days at 27 °C/180 r.p.m. Mycelia were subsequently separated from the media, washed three times and frozen in liquid nitrogen. Total RNA extraction was performed according to TRIzol-based methods (Invitrogen). For the RT reactions, total RNA was treated with DNase RQI RNase-Free (Promega) and quantified spectrophotometrically at 260 nm. To synthesize the first cDNA strand, the reaction components were mixed as follows: 5 μg total RNA, 1× RT reaction buffer (Promega), 0.5 mM dNTPs, 500 ng $(dT)_{15}$ primer (Promega), 20 units of the RNase inhibitor RNAsin (Promega) and 10 units AMV Reverse Transcriptase (Promega). The reaction mixtures were adjusted to a final volume of 20 μl and incubated at 25 °C for 5 min followed by 60 min at 42 °C. Primers CPORT1 and CPORT2 were used to amplify a fragment of the *Boto* transposase coding region. The expected size of the amplification product is 795 or 694 bp, depending on whether genomic DNA or cDNA is used as the template, respectively. The programme used for this amplification was 25 cycles of 1 min at 94 °C, 1 min at 55 °C and 1 min at 72 °C. The entire 25 μl amplification reaction was electrophoresed in a 1.5 % agarose gel. To analyse the *Boto* ORF1 expression, two primer sets, Boto2ORF1F1 (5′-AGTCTTCGGCA-ACCAATGAG) plus Boto2ORF1R1 (5′-CCTCGGGTTGGCCTT-AACATA) and Boto2ORF1F2 (5′-CAGAGCCAAACAGTGCAAAA) plus Boto2ORF1R2 (5′-CCGAGACACTCAATCCACCTG), were used. The size of the PCR product was expected to be either 402 or 347 bp and 896 or 794 bp, depending on whether genomic DNA or cDNA was used as the template, respectively. The programme used for this amplification was 35 cycles of 1 min at 94 °C, 1 min at 51 °C and 1 min at 72 °C. The reaction mixture was electrophoresed in a 2.5 % agarose gel. The amplified cDNA from *Boto* transposase and ORF1 was cloned into the pGEM-T Easy vector (Promega) according to the manufacturer's instructions. Sequencing and analysis of the cDNA were performed as described above.

RESULTS

Boto belongs to the *PIF/IS5* superfamily

Our group had previously designed a phage (λEMBL3) genomic library of *M. perniciosa* aimed at isolating complete and intact genes when only a partial gene sequence was available in the Witches' Broom Genome Project database, which was common at the beginning of the project. DNA from isolate CP02, the same isolate used in the Witches' Broom Genome Project, was used as the template for primers CPORT1 and CPORT2. A 795 bp DNA fragment, amplified by those primers and containing part of a transposase sequence, was used as a probe to screen for recombinant phages (data not shown). A 4 kb *Eco*RI fragment from phage 2.1.1 was cloned and sequenced. The resulting sequence corresponds to the element designated *Boto*. This element is 3089 bp and has TIRs of 45 bp (5′-GGGCCTGTTCGGTAAAAAAAAGCT-GTAGCTTTTTCGCAGCTTTTC and 5′-GAAAAGCTA-CGAAAAAGCTGCAGCTTTTTTTTACCGAACAGGCCC) with 95.55 % identity, varying only in two base pairs. A 3 bp sequence (TAA) was found flanking the *Boto* TIRs, thus characterizing the putative TSD. An ORF beginning 127 bp downstream of the 5′ TIR codes for a 414 aa protein, showing high similarity to transposases of plant *PIF*-like elements and hypothetical proteins of *Cryptococcus neoformans*, *Cryptococcus gattii* and *Ajellomyces capsulatus* (e-values from 2×10^{-53} to 7×10^{-46}). Therefore, although fungal *PIF*/*Harbinger* transposases have been previously described only for *C. neoformans* and *N. crassa* (Zhang *et al.*, 2001, 2004), new *PIF*/*Harbinger* transposases from two fungal species were found in GenBank. In addition to the 3 bp TSD and the similarity to *PIF*-like transposases, two other characteristics indicate that *Boto* is a member of the *PIF*/*Harbinger* superfamily: (i) the presence of two introns at the transposase coding region, and (ii) the presence of a second ORF coding for a protein of unknown function (Fig. 1).

Boto transposase contains two introns

Comparison of the *Boto* transposase with transposases of plant *PIF*/*Harbinger* elements revealed the presence of some conserved domains (Fig. 1c). These domains have already been described in transposases of plant *PIF*-like elements (Zhang *et al.*, 2004) and correspond to (i) the HTH domain (helix–turn–helix), which could participate in DNA binding, and to (ii) the N2, N3 and C1 regions that probably contain the protein catalytic domain, given that they contain the characteristic DDE amino acid residues (Asp, Asp and Glu), with one residue located in each region. The *Boto* transposase was found to have the same $DD^{48}E$ spacing reported for some transposases of plant *PIF*-like elements, which can also have the $DD^{47}E$ spacing (Zhang *et al.*, 2004). For the *Harbinger* and *Pong* transposases, this motif can be seen as $DD^{35}E$ (Kapitonov & Jurka, 1999, 2004; Zhang *et al.*, 2004).

The first intron (53 bp) in the *Boto* transposase coding region interrupts the His^{133} codon and has an A+T content of 68 %. The second intron (48 bp) interrupts the Arg^{285} codon and has an A+T content of 67 %. The transposase intron positions of *Boto* are different from

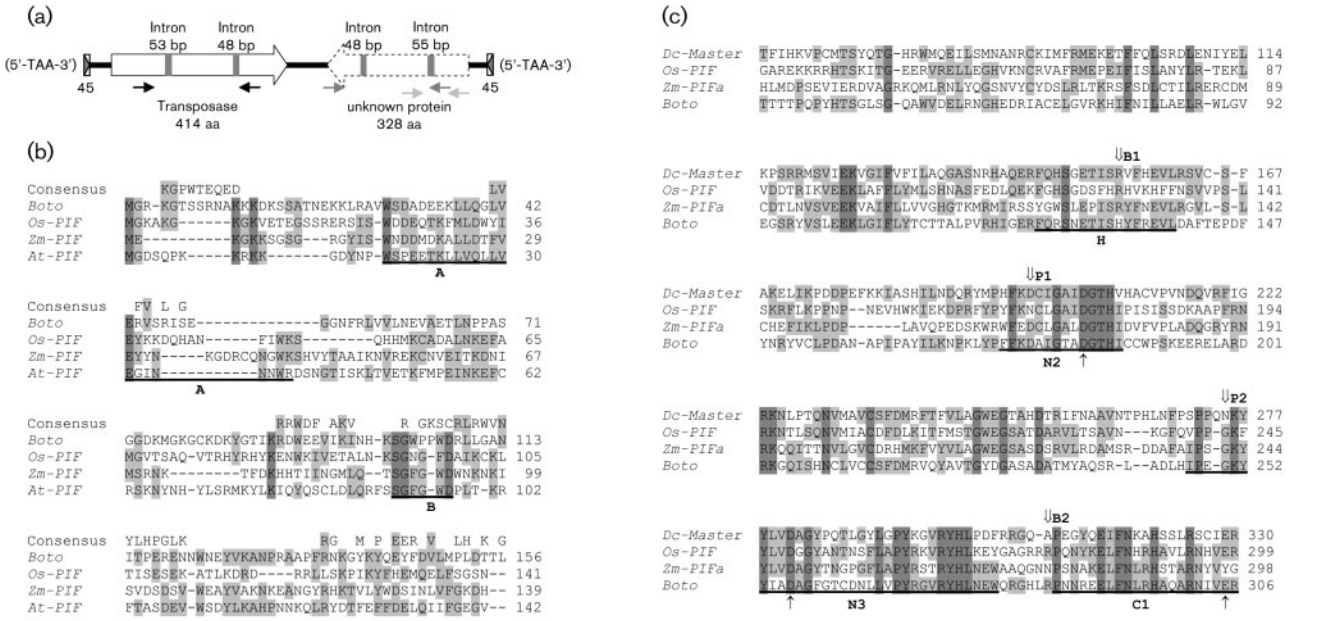

Fig. 1. (a) Schematic representation of the *Boto* element. The dotted arrow indicates the presence of ORF1 exhibiting low sequence similarity to the *myb* transcription factor. Grey boxes represent introns; small black arrows represent primers CPORT1 and CPORT2 used in the Southern blot, PCR and RT-PCR analyses; and small light and dark grey arrows represent primers Boto2ORF1F1, Boto2ORF1R1, Boto2ORF1F2 and Boto2ORF1R2 used to analyse ORF1 intron size and position. (b) Multiple alignments of the selected *PIF*/*Harbinger* ORF1. A and B indicate the two most conserved blocks identified by Zhang *et al.* (2004). 'Consensus' indicates the consensus amino acid residues obtained by the alignment of some plant *myb* transcription factors (*Oryza sativa*, AY398581; *Arabidopsis thaliana*, NM_114482; and *Glycine max*, DQ822919). The residues highlighted in grey are conserved among the analysed ORF1 proteins (*Os-PIF*, AC078977; *Zm-PIF*, EU949209; and *At-PIF*, NM_122608). (c) Multiple alignments of the *Boto* transposase protein with transposases described for plant *PIF*/*Harbinger* elements (*Os-PIF*, AAP52086; *Zm-PIFa*, AF412282; and *DC*-Master, ABB83644). Only the most conserved regions are presented. The horizontal lines indicate the HTH domain (H) and the three regions of conserved amino acids (N2, N3 and C1) that must contain the catalytic domain of the enzyme (Zhang *et al.*, 2004). The residues highlighted in grey are conserved among the analysed transposases. The DDE domain is indicated by (↑); (⇓) indicates the position of the following elements: ⇓B1, intron 1 of the *Boto* element; ⇓B2, intron 2 of the *Boto* element; ⇓P1, intron 1 of plant *PIF*-like elements; and ⇓P2, intron 2 of plant *PIF*-like elements.

those reported for plant *PIF*-like elements (Zhang *et al.*, 2004). Introns 1 and 2 in the plant *PIF*-like transposase coding region are located 6 aa residues upstream from the first and second Asp (D) of the DDE domain, respectively (Zhang *et al.*, 2004), but, in the *Boto* element, intron 1 was located 50 aa upstream from the first Asp of the DDE domain and intron 2 was located 28 aa downstream from the second Asp of the DDE domain (Fig. 1c).

Boto ORF1 also contains two introns

The sequence downstream of the transposase coding region contains a second ORF of 1090 bp, interrupted by two introns and coding for a 328 aa protein with low similarity to the DNA-binding domain of the *myb* transcription factor (Fig. 1b). Comparing that ORF with a sequence (EEB88797) presented in the Witches' Broom Genome Project Database allowed the identification of two additional thymines in the *Boto* ORF1 at positions +1067 and +1147 (based on the first ATG), which are responsible for the appearance of a premature stop codon. The removal of these additional thymines resulted in a 1372 bp ORF1 coding for a 422 aa protein, where the distance from the transposase stop codon and the ORF1 stop codon was only 16 bases. The presence of the two introns was confirmed by sequencing of PCR fragments amplified from ORF1 using cDNA and genomic DNA as templates (data not shown). These two introns are 55 bp with an A+T content of 58.2 %, and 48 bp with an A+T content of 75.0 %.

The *Boto* ORF1 intron position could not be compared with other fungal ORF1 sequences, and, although introns have been described in other *PIF*/*Harbinger* ORF1 sequences, the presence of two introns appears to be unusual. Analysis of the *Boto* ORF1 protein along with ORF1 proteins of plant *PIF*/*Harbinger* elements revealed the presence of some conserved blocks (Fig. 1b) previously identified by Zhang *et al.* (2004).

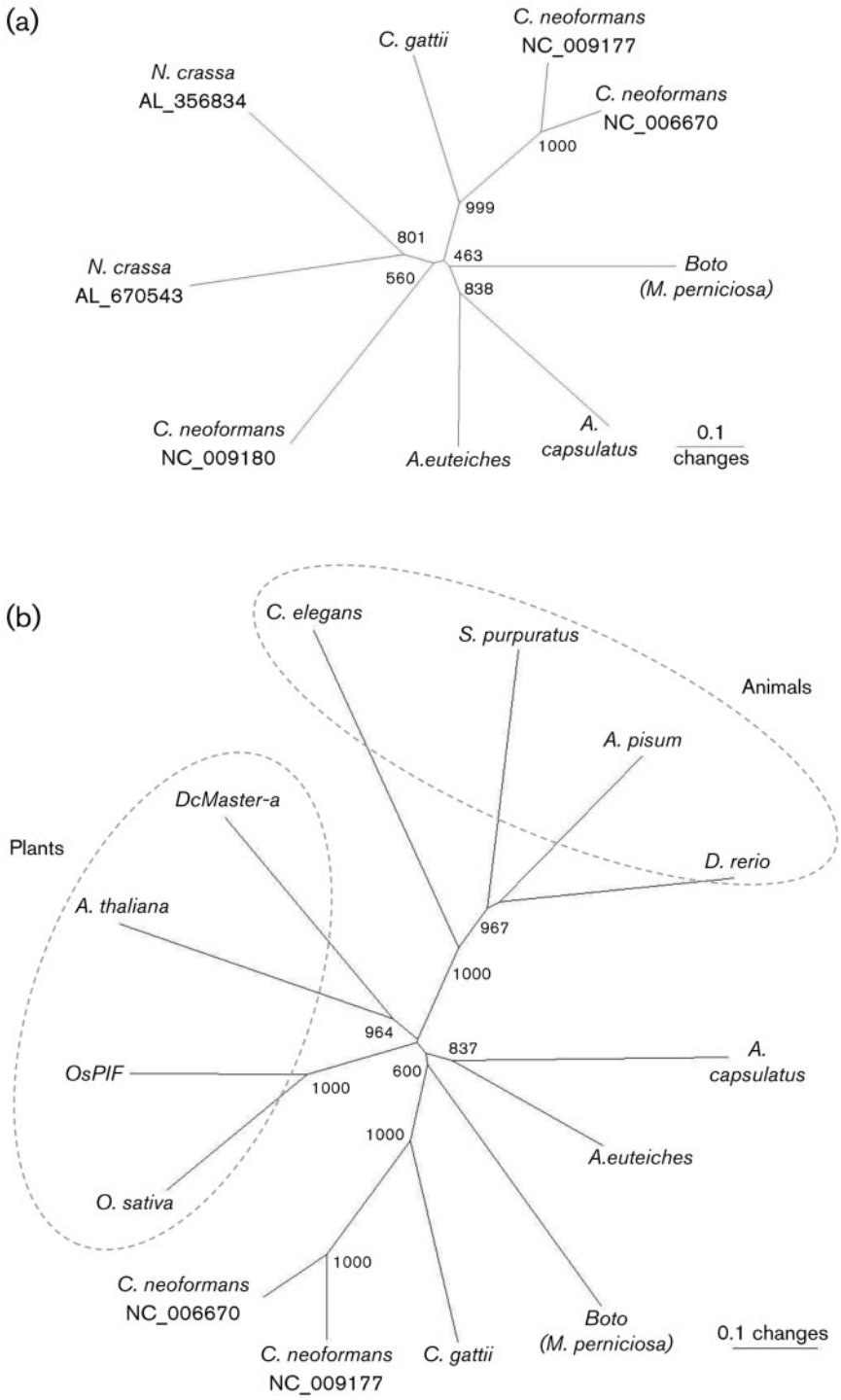

Fig. 2. Phylogenetic tree for *Boto* transposase. Trees were built with fungal and oomycete *PIF*/*Harbinger*-like transposases (a), and fungal, oomycete, plant and animal *PIF*-like tranposases (b). The trees were constructed by using the neighbour-joining method (Saitou & Nei, 1987). Numbers indicate the percentage of bootstrap replicates from a sample of 1000 that support the branches. Sequences are named according to the species or the elements. GenBank accession nos: *Acyrthosiphon pisum* (AC202214), *Ajellomyces capsulatus* (XM_001541700), *Aphanomyces euteiches* (CU363155), *Arabidopsis thaliana* (AC005850), *Boto* (EU218539), *Caenorhabditis elegans* (NM_062114), *Cryptococcus gatti* (XM_003102814), *Cryptocossus neoformans* (NC_006670, 787098–788500; NC_009177, 778738–779561; NC_009180, 174072–175467), *Danio rerio* (XM_001921333), *DcMaster-a* (DQ250806), *Neurospora crassa* (AL670543, 39714–39364; AL356834, 64784–64443), *Oryza sativa* (NM_001070615), *OsPIF* (NM_001070686) and *Strongylocentrotus purpuratus* (XM_788866).

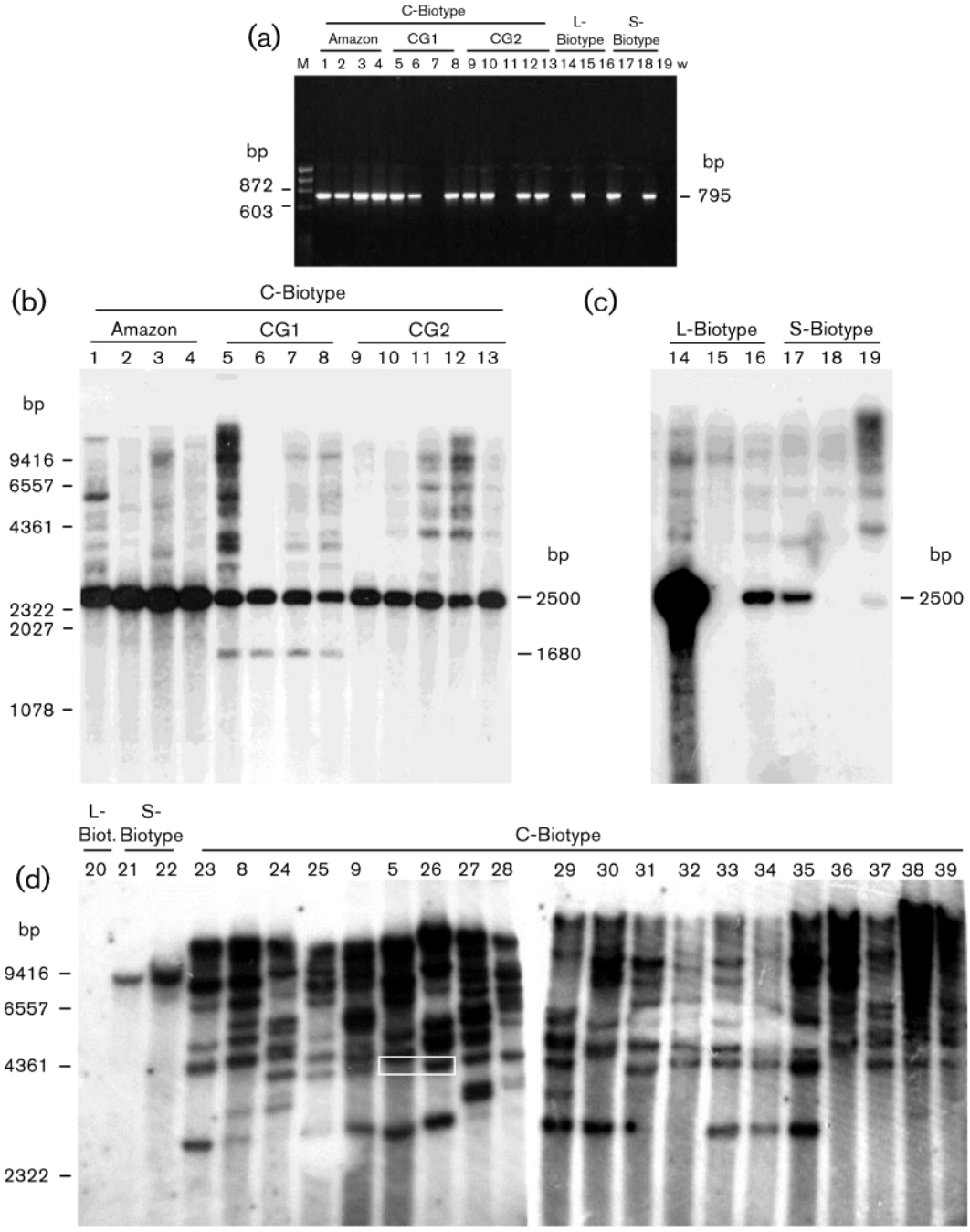

Fig. 3. PCR and hybridization analyses of the C-, S- and L-biotype isolates of *M. perniciosa*. (a) Amplification of a 795 bp fragment containing part of the *Boto* transposase coding region. (b, c) Hybridizations, performed at 58 °C, of *M. perniciosa* total DNA cleaved with *Hin*dIII. (d) Hybridizations, performed at 65 °C, of *M. perniciosa* total DNA digested with *Sal*I. White rectangle indicates the new *Boto* insertion in the CP02 isolate. In all hybridization experiments, the 795 bp *Boto* transposase fragment was used as the probe. See Table 1 for identification of isolates 1–39.

Boto and other fungal *PIF*-like transposases belong to the same phylogenetic cluster

A phylogenetic tree was constructed based on the transposase protein deduced from *Boto* and the transposases and putative proteins of fungi and an oomycete (Fig. 2a), and plants and animals (Fig. 2b). The sequences from *N. crassa* and one from *C. neoformans* (NC_009180), when analysed together with the putative transposases of plants and animals, resulted in branches with low bootstrap values (data not shown). *Boto*, the oomycete (*Aphanomyces*

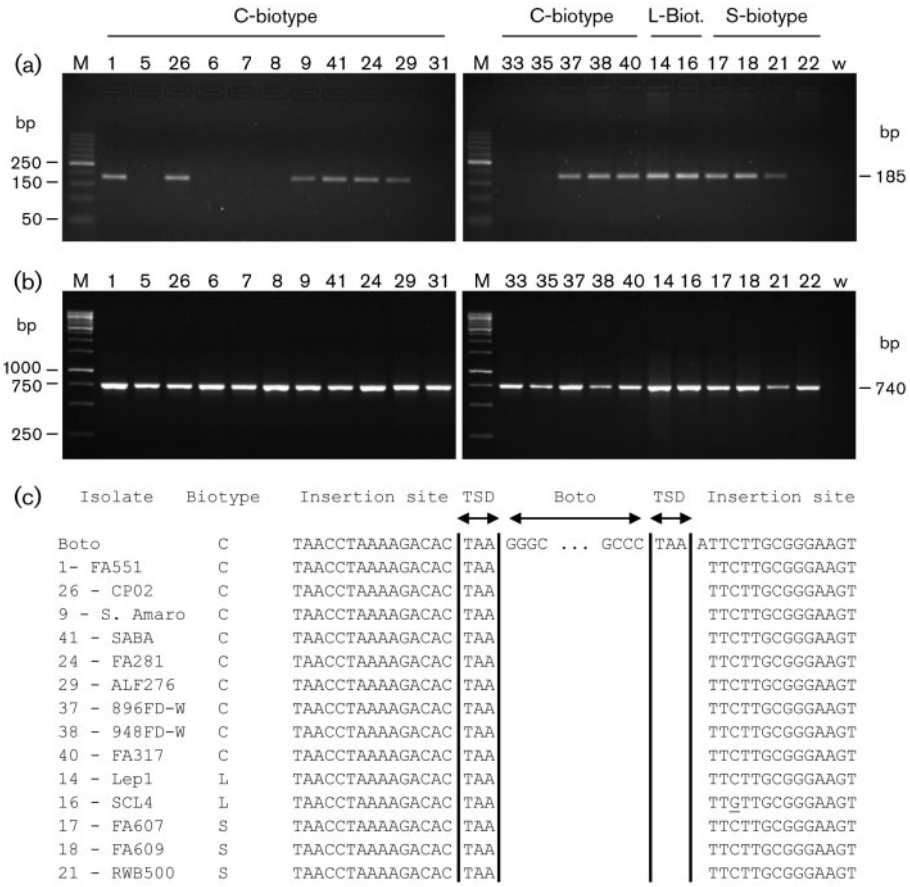

Fig. 4. (a) PCR analysis (separated on a 1.5 % agarose gel) for detection of regions without *Boto* insertion. The annealing sites for primers Boto2.1 and Boto2.2 flank the region where *Boto* was characterized, and a 185 bp PCR fragment is expected if *Boto* is not present in that region. 'M' indicates the molecular size marker (50 bp ladder). (b) ITS amplification as a DNA quality control. 'M' indicates the molecular size marker (1 kb ladder). The reaction control without DNA is denoted by 'w', and numbers indicate *M. perniciosa* isolates (see Table 1). (c) Sequence analysis for detection of *Boto* excision footprints. 'TSD' indicates the target site duplication. Underlined 'G' indicates the G : C transversion in isolate SCL4.

euteiches) protein, and the four other fungal transposases (*C. neoformans* NC_006670, *C. neoformans* NC_009177, *C. gattii* and *A. capsulatus*) grouped in the same branch (Fig. 2b).

Boto is ubiquitously distributed among *M. perniciosa*

The distribution analyses of *Boto* throughout the genomes of isolates from C-, S- and L-biotypes from the Amazon region and the states of Bahia and Minas Gerais, in Brazil, were performed by PCR and Southern hybridization (Fig. 3). In the PCR analysis, primers CPORT1 and CPORT2 were not able to amplify the 795 bp fragment in some isolates (Fig. 3a) but *Boto* was detected in those same isolates by hybridization analysis (Fig. 3b, c). This is probably due to mutations in the annealing sites of the primers. DNA quality did not interfere in the PCR analysis once we were able to amplify the internal transcribed spacer region (ITS) in all isolates (data not shown). Southern hybridization did not detect any *Boto* sequences in isolate 20, belonging to the L-biotype, and only one copy was found in the S-biotype isolates (Fig. 3d). The C-biotype strains, when analysed with the *Sal*I enzyme, exhibited from six to 12 copies of the *Boto* element (Fig. 3d). When analysed with the *Hin*dIII enzyme, the C-biotype isolates from Bahia State were divided into two groups by the presence of a 1.68 kb fragment according to their chromosomal groups (Fig. 3b). This fragment was not detected in the Amazon isolates or in the L- and S-biotype isolates analysed (Fig. 3b, c). Two different cultures of isolate CP02, the same isolate used in the Witches' Broom Genome Project, were analysed: cultures CP02 and CP02-1. Culture CP02 was maintained as a stock in tubes with water, and culture CP02-1 was successively grown in PDA medium in the laboratory for 6 years. Fig. 3(d) shows the variation in the *Boto* element profiles of these two cultures.

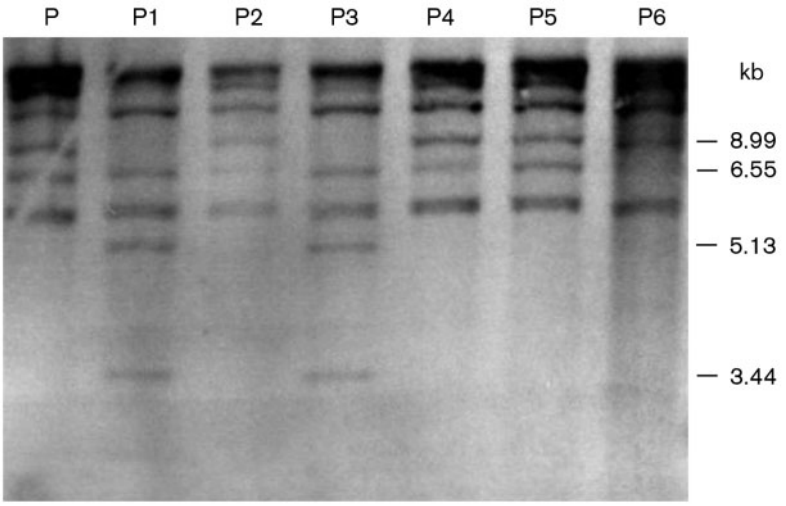

Fig. 5. Hybridization profile of an *M. perniciosa* parental isolate and six isolates from its progeny after the sexual cycle. Total DNA from *M. perniciosa* was cleaved with *Sal*I, and hybridization was performed at 65 °C. The 795 bp PCR fragment containing part of the *Boto* transposase was used as a probe. 'P' indicates the parental isolate (1919; see Table 1 for more details), and 'P1–P6' indicate the six isolates obtained from its offspring after the *M. perniciosa* sexual cycle.

Boto excision is not perfect

Based on the flanking regions of the *Boto* transposon, a primer set was designed to analyse the putative excision footprints. The extension time used to amplify that specific region was short (30 s), and was less than the time necessary for the amplification of the whole element (3089 bp). In the analyses of 22 *M. perniciosa* isolates, the expected PCR fragment (185 bp) was detected in 14 isolates belonging to C-, S- and L-biotypes (Fig. 4a). Among those isolates, the cultures CP02 (maintained in stock) and CP02-1 (cultivated in our laboratory for 6 years) were negative and positive, respectively, indicating that *Boto* transposed from that site in culture CP02-1. Sequence analysis of the 185 bp fragment in isolate CP02-1 with *Boto* flanking regions revealed that (i) the TAA is not duplicated and (ii) the first adenine, downstream of the duplicated TAA in the *Boto* element, is not present in the CP02-1 isolate (Fig. 4c). Those same characteristics were found in all of the other 13 sequences, in addition to a G : C transversion in isolate SCL4, three bases downstream of the TAA. Once *Boto* was identified through the genome sequencing of isolate CP02, analysis of isolates CP02 and CP02-1 provided evidence that *Boto* removes one copy of the 3 bp (TAA) target site duplication and just one other base during transposition. This finding is different from those reported for the *mPing* and *Harbinger* elements where a high proportion of 'perfect' excision (when, after the excision, the sequence at the insertion site is the same as before the insertion) was found (Yang *et al.*, 2007; Sinzelle *et al.*, 2008).

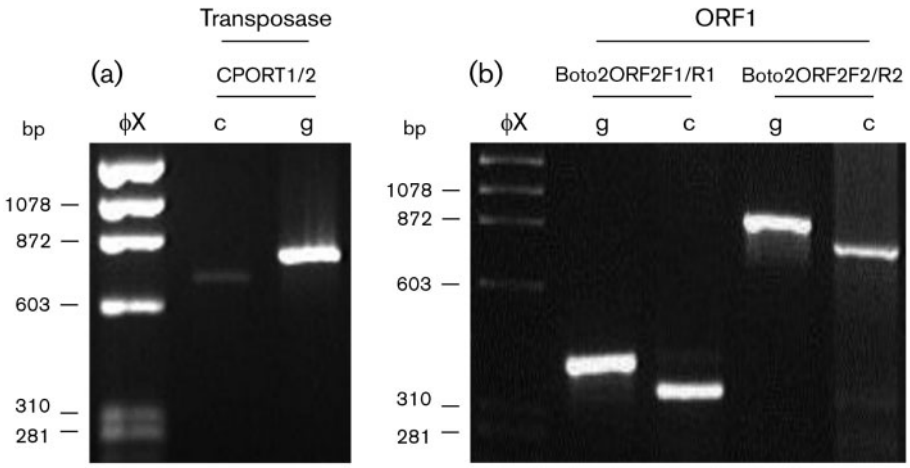

Fig. 6. Partial RT-PCR amplification of *Boto* transposase (with primers CPORT1/2) (a) and ORF1 (with primers Boto2ORF1F1/R1 and Boto2ORF1F2/R2) (b) genes using genomic DNA (g) and cDNA (c) from isolate CP02-1. The expected sizes for the PCR products are 795 or 694 bp (for primers CPORT1/2), 402 or 347 bp (for primers Boto2ORFF1/R1) and 896 or 794 bp (for primers Boto2ORFF2/R2) using genomic DNA or cDNA, respectively. 'ϕX', Molecular size marker (DNA from ϕX174 phage cleaved with *Hae*III).

A different *Boto* hybridization pattern is detected after the *M. perniciosa* sexual cycle

To analyse *Boto* transposition during the *M. perniciosa* sexual cycle, isolate 1919 was used to compare the *Boto* hybridization pattern with the pattern of its progeny (Fig. 5). Using the 795 bp *Boto* transposase fragment as a probe, six *Boto* copies were detected in the parental isolate, and a modified hybridization pattern was observed in some of its progeny. Isolates P1 and P3 had lost the 8.99 kb fragment, and two new fragments (3.44 and 5.13 kb) were detected, while isolate P6 had lost the 6.55 kb fragment. The variation in the hybridization pattern of these isolates could be explained by (i) *Boto* transposition to a new site (in isolates P1 and P3) and *Boto* excision without reinsertion (in isolate P6), (ii) recombination or (iii) a combined action of the two mechanisms.

Amplification of *Boto* transcripts

The activity of *Boto* in *M. perniciosa* was also analysed by RT-PCR (Fig. 6). RNA was extracted from a culture grown in minimal media, and *Boto* transcripts related to the transposase and ORF1 genes were amplified (Fig. 6). The sizes of the DNA fragments amplified from the cDNA were smaller than those amplified from genomic DNA, thus confirming the presence of the introns (Fig. 6).

DISCUSSION

Even though *in silico* analysis has revealed a higher number of class I than class II transposable elements in *M. perniciosa* (Mondego *et al.*, 2008), we were able to identify, at the start of the Witches' Broom Genome Project, a rare transposase sequence for a phytopathogenic fungus. This paper describes the complete characterization of that class II element, named *Boto*, in the *M. perniciosa* genome. Some characteristics of the *Boto* transposon are similar to those of *PIF*/*Harbinger* elements, including: (i) 3 bp (TAA) target site duplication; (ii) small TIRs (45 bp); (iii) a second ORF (at the −2 frame) that codes for a protein exhibiting low sequence similarity to the plant *myb* transcription factor; (iv) sequence similarity of *Boto* transposase as well as the $DD^{48}E$ spacing; and (v) two introns at the transposase coding region. *PIF*-like and *Pong*-like elements of plants also have a 3 bp TSD, which is usually TTA or TAA, although the *PIF*-like element TSD was characterized as AAT in *C. neoformans* (Zhang *et al.*, 2001). A 3 bp TSD was also characterized in all autonomous and non-autonomous *Harbinger* elements analysed by Kapitonov & Jurka (2004). Although the length of TIRs for *Harbinger* elements may vary from 10 to 700 bp (Kapitonov & Jurka, 2004), the size varies from 10 to 45 bp in the *PIF*-like elements of rice (Zhang *et al.*, 2004), similar to the 45 bp size determined for the *Boto* element TIRs. As described for TIRs from most *OsPIFs* and *OsPongs* elements, the *Boto* TIRs also begin with 5′-GGSG-3′ (where S represents G or C). Specific *PIF* or *Pong* inner TIR motifs were not identified, although bases 6–14 in the *Boto* TIRs (5′-TGTTCGGTA-3′) are more similar to *PIF* (5′-TGTTTGGTT-3′) than *Pong* elements (Zhang *et al.*, 2004). ORF1 exhibits weak similarity to transcription factors that may have a possible role in the transposition mechanism (Kapitonov & Jurka, 2004; Zhang *et al.*, 2004; Yang *et al.*, 2007; Sinzelle *et al.*, 2008; Hancock *et al.*, 2010). Assuming a role of ORF1 in transposition, it is possible that *Boto* transposition could be achieved by cross-mobilization if the *Boto* ORF1 protein fails to produce a functional protein due to the presence of two additional thymines in its coding region.

The *PIF*/*Harbinger* elements are not abundantly distributed in fungal genomes. This fact is not a reflection of the number of fungal genomes currently available because close to 500 genomes, including yeasts, are sequenced or near completion (Keyhani, 2011). At least two hypotheses can explain the low distribution of the *PIF*/*Harbinger* elements in fungal genomes: (i) these elements have been lost during evolution in the majority of the fungal species studied so far or (ii) horizontal transfer spreads these elements to only some fungal genomes. The hypothesis of horizontal transfer is supported by the sporadic and non-homogeneous distribution of *PIF*/*Harbinger* transposases observed in fungi, having only been detected in three human-pathogenic fungi (*C. neoformans*, *C. gattii* and *A. capsulatus*), one saprotrophic fungus (*N. crassa*) and one phytopathogen (*M. perniciosa*). Such non-uniform distribution of an element within isolates of a single species (or within the same group, as in the present case) may reflect the recent acquisition of this element (Daboussi & Capy, 2003). Horizontal transfer was hypothesized to have a role in the distribution of *Harbinger* transposons in plants (Kapitonov & Jurka, 2004) and in some fungal transposons from different classes and superfamilies (Dobinson *et al.*, 1993; Daboussi & Langin, 1994; He *et al.*, 1996; Shull & Hamer, 1996; Nakayashiki *et al.*, 1999; Shim & Dunkle, 2005). Although in the phylogenetic analysis of *Boto*, one oomycete and four fungal transposases grouped in the same branch (Fig. 2b), Zhang *et al.* (2004) reported that *PIF*-like transposases in *C. neoformans* and *N. crassa* formed two distinct species-specific groups that failed to show a common ancestor when analysed with 600 other *PIF*-like transposases of plants and animals. Undoubtedly, more detailed analyses are necessary to elucidate the polyphyletic nature of fungal *PIF*-like transposases, but the role of horizontal transfer cannot be ruled out.

MITEs are present in high copy numbers in plant genomes (Wessler *et al.*, 1995). Given that *PIF*/*Harbinger* elements are present in several plant species and are directly linked to the origin and mobilization of MITEs in plants (Zhang *et al.*, 2001, 2004; Kapitonov & Jurka, 2004; Grzebelus *et al.*, 2006), the wide distribution of MITEs in plants can be expected. Therefore, the low distribution of *PIF*/*Harbinger* elements in fungal genomes may be related to the small number of MITE-like elements found in filamentous fungi (Yeadon & Catcheside, 1995; Hua-Van *et al.*, 2000; Fleetwood *et al.*, 2007, 2011). Although the *PIF*/*Harbinger*

elements have been identified as sources of transposases for *Tourist*-like MITEs in maize and rice (Zhang *et al.*, 2001; Jiang *et al.*, 2003), the cross-mobilization of the *mimp* elements from *Fusarium oxysporum* was linked to the *impala* transposase, an element of the *TC1/Mariner* superfamily (Dufresne *et al.*, 2007; Bergemann *et al.*, 2008).

In *M. perniciosa*, *Boto* elements were found to be ubiquitous among the analysed isolates belonging to the C-, L- and S-biotypes (Fig. 3). Although a 2.5 kb *Hin*dIII DNA fragment is conserved in the *M. perniciosa* isolates analysed (Fig. 3b, c), a 1.68 kb *Hin*dIII fragment did distinguish the C-biotypes from Bahia State, the major state of cacao production in Brazil, into two different groups related to the chromosomal groups described by Rincones *et al.* (2006). Genetic variability studies in *M. perniciosa*, using several different molecular techniques, have revealed two different genotypic groups in Bahia State and genetically close relationships between a number of isolates from that state with isolates from the Amazon region. Those data have been used to propose (Andebrhan *et al.*, 1999) and to corroborate the hypothesis (de Arruda *et al.*, 2003a, b; Rincones *et al.*, 2003) that the witches' broom outbreak in Bahia State occurred by two independent focal points of introduction. Moreover, a reverse transcriptase sequence, part of a putative *gypsy*-like retrotransposon, and a transposase sequence, belonging to the *TC1-Mariner* superfamily, also distinguished C-biotype isolates from Bahia in two genotypic groups (Pereira *et al.*, 2007; Ignacchiti *et al.*, 2011). Thus, some transposable elements appear to spread through the *M. perniciosa* genome in accordance with some chromosomal groups.

Different strategies could be used to demonstrate transposon activity such as gene inactivation (Daboussi *et al.*, 1992; Langin *et al.*, 1995; Maurer *et al.*, 1997; Gómez-Gómez *et al.*, 1999; Ogasawara *et al.*, 2009), detection of new insertions in the genome (Anaya & Roncero, 1996; Mes *et al.*, 2000) or through expression analyses (Okuda *et al.*, 1998; Kaneko *et al.*, 2000; Kito *et al.*, 2003; Rep *et al.*, 2005; Ogasawara *et al.*, 2009). For *Boto* elements of *M. perniciosa*, the last two strategies were used to give experimental support to their expression and activity, including (i) variation found in the hybridization profiles in different cultures of isolate CP02 (Fig. 3d); (ii) PCR amplification from the specific site where *Boto* was characterized (Fig. 4); (iii) different hybridization patterns in three isolates originated after the *M. perniciosa* sexual cycle (Fig. 5); (iv) successful amplification of *Boto* transcripts (Fig. 6); and (v) variation in the copy number and location in different isolates of this fungus (Fig. 3). Thus, we conclude that the *Boto* family is active and may contribute to the genetic variability in *M. perniciosa*.

One possible explanation for the *Boto* activity observed when *M. perniciosa* was subjected to the sexual cycle is based on the idea that transposable elements could be activated under stress conditions. That idea is supported by the fact that transposition contributes to the generation of genetic variability, which could confer adaptive advantages to the organism under environmental stress (McClintock, 1984). Other genes involved in transposition and retrotransposition appear to be activated during the *M. perniciosa* life cycle and were found among the 189 genes that showed significantly different expression between biotrophic-like and saprotrophic mycelia (Rincones *et al.*, 2008). Moreover, transposition activity is not necessary for the mutagenic effects of transposable elements, as the copies throughout the genome can be used for reorganization through ectopic recombination (Daboussi & Capy, 2003; Shnyreva, 2003).

The presence of the two introns in the *Boto* transposase coding region was confirmed by RT-PCR. Sequencing data revealed that these introns are small and show a high A + T content, thus resembling the introns found in plant *PIF*-like elements. However, the intron position found in the *Boto* element differs from that reported for plant *PIF*-like elements. The first intron (53 bp) is located 50 aa residues upstream of the first Asp (D) of the DDE domain, and the second intron (48 bp) is located 28 aa residues downstream of the second Asp (D) of the DDE domain. In *PIF/Harbinger* elements, different arrangements are found for the ORFs that code for the transposase and the protein of unknown function (Kapitonov & Jurka, 2004; Zhang *et al.*, 2004). These ORFs may be oriented in the same or in opposite directions, and the transposase ORF can be found upstream or downstream of the unknown protein ORF. The arrangement found in the *Boto* transposon (ORFs in opposite directions and the transposase ORF upstream of the unknown protein ORF) has not been described for plant *PIF/Harbinger* elements (Zhang *et al.*, 2004) but is similar to the arrangements found in some families of *Harbinger* elements in animals (*Anopheles gambiae* and zebrafish) and a protist (*Thalassiosira pseudonana*) (Kapitonov & Jurka, 2004).

Considering our results, the *Boto* element of *M. perniciosa* has evolved differently from previously described *PIF/Harbinger* elements, and a few differences are thus expected between the transposases of these elements. Given the particular characteristics with regard to transposase intron position, the organization of the second ORF and the footprint, the *Boto* element of *M. perniciosa* belongs to a new family of transposable elements of the *PIF/Harbinger* superfamily. This is an active family of transposable elements in *M. perniciosa* that may contribute to the genome plasticity and adaptability of this phytopathogenic fungus.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Brazilian agencies CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and SEAGRI (Secretaria de Agricultura do Estado da Bahia) for financial support. Dr A. Leyva helped with the English editing of the manuscript.

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Anaya, N. & Roncero, M. I. G. (1996). Stress-induced rearrangement of *Fusarium* retrotransposon sequences. *Mol Gen Genet* **253**, 89–94.

Andebrhan, T. & Furtek, D. B. (1994). Random amplified polymorphic DNA (RAPD) analysis of *Crinipellis perniciosa* isolates from different hosts. *Plant Pathol* **43**, 1020–1027.

Andebrhan, T., Figueira, A., Yamada, M. M., Cascardo, J. & Furtek, D. B. (1999). Molecular fingerprinting suggests two primary outbreaks of witches' broom disease (*Crinipellis perniciosa*) of *Theobroma cacao* in Bahia, Brazil. *Eur J Plant Pathol* **105**, 167–175.

Bastos, C. N. & Evans, H. C. (1985). A new pathotype of *Crinipellis perniciosa* (Witches' broom disease) on solanaceous hosts. *Plant Pathol* **34**, 306–312.

Bastos, C. N., Andebrhan, T. & de Almeida, L. C. (1988). Comparação morfológica de isolados de *Crinipellis perniciosa*. *Fitopatol Bras* **13**, 202–205.

Benton, W. D. & Davis, R. W. (1977). Screening lambdagt recombinant clones by hybridization to single plaques in situ. *Science* **196**, 180–182.

Bergemann, M., Lespinet, O., M'Barek, S. B., Daboussi, M. J. & Dufresne, M. (2008). Genome-wide analysis of the *Fusarium oxysporum mimp* family of MITEs and mobilization of both native and de novo created *mimps*. *J Mol Evol* **67**, 631–642.

Daboussi, M. J. & Capy, P. (2003). Transposable elements in filamentous fungi. *Annu Rev Microbiol* **57**, 275–299.

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

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

de Arruda, M. C., Ferreira, M. A., Miller, R. N., Resende, M. L. & Felipe, M. S. (2003a). Nuclear and mitochondrial rDNA variability in *Crinipellis perniciosa* from different geographic origins and hosts. *Mycol Res* **107**, 25–37.

de Arruda, M. C., Miller, R. N., Ferreira, M. A. & Felipe, M. S. (2003b). Comparison of *Crinipellis perniciosa* isolates from Brazil by ERIC repetitive element sequence-based PCR genomic fingerprint. *Plant Pathol* **52**, 236–244.

Dobinson, K. F., Harris, R. E. & Hamer, J. E. (1993). Grasshopper, a long terminal repeat (LTR) retroelement in the phytopathogenic fungus *Magnaporthe grisea*. *Mol Plant Microbe Interact* **6**, 114–126.

Dufresne, M., Hua-Van, A., El Wahab, H. A., Ben M'Barek, S., Vasnier, C., Teysset, L., Kema, G. H. J. & Daboussi, M. J. (2007). Transposition of a fungal miniature inverted-repeat transposable element through the action of a *Tc1*-like transposase. *Genetics* **175**, 441–452.

Evans, H. C. (1978). Witches' broom disease of cocoa (*Crinipellis perniciosa*) in Ecuador. *Ann Appl Biol* **89**, 185–192.

Felipe, M. S. S., Azevedo, M. A., Vainstein, M. H. & Schrank, A. (1992). Biologia molecular de fungos filamentosos: construção de banco genômico e de cDNA. In *Manual Técnico*, p. 99. Piracicaba, SP, Brazil: Escola Superior de Agricultura 'Luiz de Queiroz'.

Fleetwood, D. J., Scott, B., Lane, G. A., Tanaka, A. & Johnson, R. D. (2007). A complex ergovaline gene cluster in epichloe endophytes of grasses. *Appl Environ Microbiol* **73**, 2571–2579.

Fleetwood, D. J., Khan, A. K., Johnson, R. D., Young, C. A., Mittal, S., Wrenn, R. E., Hesse, U., Foster, S. J., Schardl, C. L. & Scott, B. (2011). Abundant degenerate miniature inverted-repeat transposable elements in genomes of epichloid fungal endophytes of grasses. *Genome Biol Evol* **3**, 1253–1264.

Gómez-Gómez, E., Anaya, N., Roncero, M. I. G. & Hera, C. (1999). *Folyt1*, a new member of the hAT family, is active in the genome of the plant pathogen *Fusarium oxysporum*. *Fungal Genet Biol* **27**, 67–76.

Goodwin, T. J. & Poulter, R. T. (2001). The DIRS1 group of retrotransposons. *Mol Biol Evol* **18**, 2067–2082.

Goodwin, T. J., Butler, M. I. & Poulter, R. T. (2003). Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* **149**, 3099–3109.

Griffith, G. W. & Hedger, J. N. (1993). A novel method for producing basidiocarps of the cocoa pathogen *Crinipellis perniciosa* using a bran-vermiculite medium. *Eur J Plant Pathol* **99**, 227–230.

Grzebelus, D., Yau, Y. Y. & Simon, P. W. (2006). *Master*: a novel family of *PIF*/*Harbinger*-like transposable elements identified in carrot (*Daucus carota* L.). *Mol Genet Genomics* **275**, 450–459.

Hancock, C. N., Zhang, F. & Wessler, S. R. (2010). Transposition of the *Tourist*-MITE *mPing* in yeast: an assay that retains key features of catalysis by the class 2 *PIF*/*Harbinger* superfamily. *Mob DNA* **1**, 5.

He, C., Nourse, J. P., Irwin, J. A. G., Manners, J. M. & Kelemu, S. (1996). *Cg*T1: a non-LTR retrotransposon with restricted distribution in the fungal phytopathogen *Colletotrichum gloeosporioides*. *Mol Gen Genet* **252**, 320–331.

Hedger, J. N., Pickering, V. & Aragundi, J. (1987). Variability of populations of the witches' broom disease of cocoa (*Crinipellis perniciosa*). *Trans Br Mycol Soc* **88**, 533–546.

Hua-Van, A., Davière, J. M., Kaper, F., Langin, T. & Daboussi, M. J. (2000). Genome organization in *Fusarium oxysporum*: clusters of class II transposons. *Curr Genet* **37**, 339–347.

Ignacchiti, M. D. C., Santana, M. F., Araújo, E. F. & Queiroz, M. V. (2011). The distribution of a transposase sequence in *Moniliophthora perniciosa* confirms the occurrence of two genotypes in Bahia, Brazil. *Trop Plant Pathol* **36**, 276–286.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S. R., McCouch, S. R. & Wessler, S. R. (2003). An active DNA transposon family in rice. *Nature* **421**, 163–167.

Jurka, J. & Kapitonov, V. V. (2001). *PIFs* meet *Tourists* and *Harbingers*: a superfamily reunion. *Proc Natl Acad Sci U S A* **98**, 12315–12316.

Kaneko, I., Tanaka, A. & Tsuge, T. (2000). *REAL*, an LTR retrotransposon from the plant pathogenic fungus *Alternaria alternata*. *Mol Gen Genet* **263**, 625–634.

Kapitonov, V. V. & Jurka, J. (1999). Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* **107**, 27–37.

Kapitonov, V. V. & Jurka, J. (2004). *Harbinger* transposons and an ancient *HARBI1* gene derived from a transposase. *DNA Cell Biol* **23**, 311–324.

Keyhani, N. O. (2011). Fungal genomes and beyond. *Fungal Genom Biol* **1**, e101.

Kito, H., Takahashi, Y., Sato, J., Fukiya, S., Sone, T. & Tomita, F. (2003). *Occan*, a novel transposon in the *Fot1* family, is ubiquitously found in several *Magnaporthe grisea* isolates. *Curr Genet* **42**, 322–331.

Langin, T., Capy, P. & Daboussi, M. J. (1995). The transposable element *impala*, a fungal member of the *Tc1-mariner* superfamily. *Mol Gen Genet* **246**, 19–28.

Marchler-Bauer, A. & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32** (Web Server issue), W327–W331.

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 of the nitrate reductase structural gene. *Mol Gen Genet* **256**, 195–202.

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* **226**, 792–801.

Mes, J. J., Haring, M. A. & Cornelissen, B. J. C. (2000). *Foxy*: an active family of short interspersed nuclear elements from *Fusarium oxysporum*. *Mol Gen Genet* **263**, 271–280.

Mondego, J. M. C., Carazzolle, M. F., Costa, G. G. L., Formighieri, E. F., Parizzi, L. P., Rincones, J., Cotomacci, C., Carraro, D. M., Cunha, A. F. & other authors (2008). A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao. *BMC Genomics* **9**, 548.

Nakayashiki, H., Nishimoto, N., Ikeda, K., Tosa, Y. & Mayama, S. (1999). Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. *Mol Gen Genet* **261**, 958–966.

Niella, G., Resende, M. L., Castro, H. A., de Carvalho, G. A. & Silva, L. H. C. P. (1999). Aperfeiçoamento da metodologia de produção artificial de basidiocarpos de *Crinipellis perniciosa*. *Fitopatol Bras* **24**, 523–527.

Ogasawara, H., Obata, H., Hata, Y., Takahashi, S. & Gomi, K. (2009). *Crawler*, a novel *Tc1/mariner*-type transposable element in *Aspergillus oryzae* transposes under stress conditions. *Fungal Genet Biol* **46**, 441–449.

Okuda, M., Ikeda, K., Namiki, F., Nishi, K. & Tsuge, T. (1998). *Tfo1*: an *Ac*-like transposon from the plant pathogenic fungus *Fusarium oxysporum*. *Mol Gen Genet* **258**, 599–607.

Pereira, J. L., de Almeida, L. C. C. & Santos, S. M. (1996). Witches' broom disease of cocoa in Bahia: attempts at eradication and containment. *Crop Prot* **15**, 743–752.

Pereira, J. F., Araújo, E. F., Brommonschenkel, S. H. & Queiroz, M. V. (2006). Elementos transponíveis em fungos fitopatogênicos. *Rev Anual Patol Plant* **14**, 303–362.

Pereira, J. F., Ignacchiti, M. D. C., Araújo, E. F., Brommonschenkel, S. H., Cascardo, J. C. M., Pereira, G. A. G. & Queiroz, M. V. (2007). PCR amplification and sequence analyses of reverse transcriptase-like genes in *Crinipellis perniciosa* isolates. *Fitopatol Bras* **32**, 373–380.

Ploetz, R. C., Schnell, R. J., Ying, Z. T., Zheng, Q., Olano, C. T., Motamayor, J. C. & Johnson, E. S. (2005). Analysis of molecular diversity in *Crinipellis perniciosa* with AFLP markers. *Eur J Plant Pathol* **111**, 317–326.

Pontecorvo, G., Roper, J. A. L., Hemmons, L. M., MacDonald, K. D. & Bufton, A. W. J. (1953). The genetics of *Aspergillus nidulans*. *Adv Genet* **5**, 141–238.

Rep, M., van der Does, H. C. & Cornelissen, B. J. (2005). *Drifter*, a novel, low copy hAT-like transposon in *Fusarium oxysporum* is activated during starvation. *Fungal Genet Biol* **42**, 546–553.

Rincones, J., Meinhardt, L. W., Vidal, B. C. & Pereira, G. A. G. (2003). Electrophoretic karyotype analysis of *Crinipellis perniciosa*, the causal agent of witches' broom disease of *Theobroma cacao*. *Mycol Res* **107**, 452–458.

Rincones, J., Mazotti, G. D., Griffith, G. W., Pomela, A., Figueira, A., Leal, G. A., Jr, Queiroz, M. V., Pereira, J. F., Azevedo, R. A. & other authors (2006). Genetic variability and chromosome-length polymorphisms of the witches' broom pathogen *Crinipellis perniciosa* from various plant hosts in South America. *Mycol Res* **110**, 821–832.

Rincones, J., Scarpari, L. M., Carazzolle, M. F., Mondego, J. M. C., Formighieri, E. F., Barau, J. G., Costa, G. G. L., Carraro, D. M., Brentani, H. P. & other authors (2008). Differential gene expression between the biotrophic-like and saprotrophic mycelia of the witches' broom pathogen *Moniliophthora perniciosa*. *Mol Plant Microbe Interact* **21**, 891–908.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463–5467.

Schmidt, S. M. & Panstruga, R. (2011). Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? *Curr Opin Plant Biol* **14**, 392–399.

Shim, W. B. & Dunkle, L. D. (2005). Malazy, a degenerate, species-specific transposable element in *Cercospora zeae-maydis*. *Mycologia* **97**, 349–355.

Shnyreva, A. V. (2003). Transposable elements are the factors involved in various rearrangements and modifications of fungal genomes. *Russ J Genet* **39**, 505–518.

Shull, V. & Hamer, J. E. (1996). Genetic differentiation in the rice blast fungus revealed by the distribution of the *Fosbury* retrotransposon. *Fungal Genet Biol* **20**, 59–69.

Sinzelle, L., Kapitonov, V. V., Grzela, D. P., Jursch, T., Jurka, J., Izsvák, Z. & Ivics, Z. (2008). Transposition of a reconstructed *Harbinger* element in human cells and functional homology with two transposon-derived cellular genes. *Proc Natl Acad Sci U S A* **105**, 4715–4720.

Specht, C. A., DiRusso, C. C., Novotny, C. P. & Ullrich, R. C. (1982). A method for extracting high-molecular-weight deoxyribonucleic acid from fungi. *Anal Biochem* **119**, 158–163.

Stanke, M. & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* **33** (Web Server issue), W465–W467.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Walker, E. L., Eggleston, W. B., Demopulos, D., Kermicle, J. & Dellaporta, S. L. (1997). Insertions of a novel class of transposable elements with a strong target site preference at the *r* locus of maize. *Genetics* **146**, 681–693.

Wessler, S. R., Bureau, T. E. & White, S. E. (1995). LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev* **5**, 814–821.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M. & other authors (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**, 973–982.

Wöstemeyer, J. & Kreibich, A. (2002). Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Curr Genet* **41**, 189–198.

Yang, G. J., Zhang, F., Hancock, C. N. & Wessler, S. R. (2007). Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **104**, 10962–10967.

Yeadon, P. J. & Catcheside, D. E. (1995). *Guest*: a 98 bp inverted repeat transposable element in *Neurospora crassa*. *Mol Gen Genet* **247**, 105–109.

Zhang, X., Feschotte, C., Zhang, Q., Jiang, N., Eggleston, W. B. & Wessler, S. R. (2001). *P* instability factor: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc Natl Acad Sci U S A* **98**, 12572–12577.

Zhang, X., Jiang, N., Feschotte, C. & Wessler, S. R. (2004). *PIF*- and *Pong*-like transposable elements: distribution, evolution and relationship with *Tourist*-like miniature inverted-repeat transposable elements. *Genetics* **166**, 971–986.

Zhou, M. B., Lu, J.-J., Zhong, H., Liu, X.-M. & Tang, D.-Q. (2010). Distribution and diversity of *PIF*-like transposable elements in the Bambusoideae subfamily. *Plant Sci* **179**, 257–266.

Zhou, M. B., Liu, X. M. & Tang, D. Q. (2012). *PpPIF-1*: first isolated full-length *PIF*-like element from the bamboo *Phyllostachys pubescens*. *Genet Mol Res* **11**, 810–820.

Edited by: R. P. Oliver