**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 Pereira5 and Marisa Vieira de Queiroz1

1Universidade Federal de Viçosa, Departamento de Microbiologia, CEP 36571-000, Viçosa, MG, Brazil
2Universidade Estadual de Maringá, Departamento de Biologia Celular e Genética, CEP 87020-900, Maringá, PR, Brazil
3CEPLAC, Centro de Pesquisa do Cacau, CEP 45600-000, Itabuna, BA, Brazil
4Universidade Federal de Viçosa, Departamento de Fitopatologia, CEP 36571-000, Viçosa, MG, Brazil
5Universidade Estadual de Campinas, Departamento de Genética e Evolução, CEP 13083-970, Campinas, SP, Brazil

**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 \textit{Tc1/mariner}, \textit{hAT}, \textit{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 \textit{et al.}, 2003).

One of the 10 class II superfamilies identified so far in eukaryotic organisms is the \textit{PIF}/\textit{Harbinger} superfamily. The first two elements described in this superfamily were the \textit{PIF} element (P instability factor) of maize (Walker \textit{et al.}, 1997) and the \textit{Harbinger} element of \textit{Arabidopsis thaliana} (Kapitonov & Jurka, 1999). \textit{PIF}/\textit{Harbinger} elements share characteristics with other groups of transposons, such as the small TIRs and the 3 bp TSD. However, some unique characteristics distinguish \textit{PIF}/\textit{Harbinger} elements from other superfamilies: (i) the presence of a transposase and the other for a protein of unknown function but showing weak similarity to \textit{myb} transcription factors (Jiang \textit{et al.}, 2003); (ii) a distant relationship between the \textit{PIF}/\textit{Harbinger} transposase and the transposase of bacterial insertion sequences (IS) of the IS5 group; and (iii) their direct link in origin and mobility of non-autonomous MITEs (Zhang \textit{et al.}, 2001, 2004; Grzebelus \textit{et al.}, 2006). The \textit{Harbinger} and \textit{PIF} elements, in addition to the rice element named \textit{Pong} (Zhang \textit{et al.}, 2004), can be seen as the founding members of this widespread superfamily of DNA transposons. A distribution analysis identified more than 600 \textit{PIF}-like transposases from 35 species of plants and 19 species of animals (Zhang \textit{et al.}, 2004), and different \textit{PIF}/\textit{Harbinger} families have been found in protists, plants, insects, worms and vertebrates (Jurka & Kapitonov, 2001; Kapitonov & Jurka, 2004; Grzebelus \textit{et al.}, 2006; Zhou \textit{et al.}, 2010, 2012). Curiously, sequences similar to \textit{PIF}-like elements were reported in only two species of fungi, \textit{Cryptococcus neoformans} and \textit{Neurospora crassa} (Zhang \textit{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 \textit{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 \textit{et al.}, 2006; Schmidt & Pantruga, 2011). In this context, studying transposable elements in the plant pathogen \textit{Moniliophthora} (formerly \textit{Criptocellis}) \textit{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 \textit{et al.}, 1996). In addition to cacao (\textit{Theobroma cacao}), \textit{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 \textit{et al.}, 1988), the S-biotype infects plants of the family Solanaceae (Bastos & Evans, 1983; Bastos \textit{et al.}, 1988) and the L-biotype is a saprotroph that colonizes a wide variety of substrates (Evans, 1978; Hedger \textit{et al.}, 1987). The genetic variability of \textit{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 \textit{et al.}, 1999; de Arruda \textit{et al.}, 2003a, b; Rincones \textit{et al.}, 2003, 2006; Ploetz \textit{et al.}, 2005).

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

**METHODS**

**Fungal strains and growth conditions.** Isolates of \textit{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 \textit{et al.} (1999).

**Isolation of recombinant phages.** A sequence showing similarity to plant \textit{PIF}-like transposase (e-value 1 $\times 10^{-28}$) was obtained from the database of the Witches’ Broom Genome Project. Primers CPORT1 (5’-TTGCTTGTGACTTGTGTC) and CPORT2 (5’-GCTGAGCATGTCGAAGATT) 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 \textit{M. perniciosa} cloned into the \textit{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 \textit{et al.} (1992). Cleavage of the phage DNA was performed using different restriction enzymes according to the manufacturer’s instructions.

**Cloning of the \textit{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 dye deoxyribonucleotide chain-termination method (Sanger \textit{et al.}, 1977) in a MegaBACE 500 sequencer (Amersham Biosciences). Analyses of DNA and protein sequences were performed using the \textit{blast} algorithm (Altschul \textit{et al.}, 1997), \textit{CLUSTAL W} program (Thompson \textit{et al.}, 1994), the CD-Search program to identify conserved domains (Marchler-Bauer & Bryant, 2004) and the \textit{AUGUSTUS} program for gene prediction (Stanke & Morgenstern, 2005).

**Phylogenetic analysis.** The sequences of the fungal, oomycete, plant and animal \textit{PIF}/\textit{Harbinger} transposase proteins were obtained from GenBank. The sequences were aligned using the \textit{CLUSTAL W} program, and phylogenetic analyses were performed based on the
neighbour-joining method (Saitou & Nei, 1987) using bootstrap values based on 1000 replicates.

**Footprint analysis.** Primers Boto2.1 (5'-TGTAGGCATTCCGAATTCGTCG) and Boto2.2 (5'-TCCGATGCTCTTGGCCGT) were designed based on the Boto flanking regions present in the δ phase 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

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

<table>
<thead>
<tr>
<th>Isolate number</th>
<th>Isolate identification</th>
<th>Biotype</th>
<th>Chromosomal group*</th>
<th>Location†</th>
<th>Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FA551</td>
<td>C</td>
<td>–</td>
<td>Tabatinga/AM</td>
<td>Theobroma sp.</td>
</tr>
<tr>
<td>2</td>
<td>ESJ0H1</td>
<td>C</td>
<td>–</td>
<td>Marituba/PA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>3</td>
<td>ESJ0H2</td>
<td>C</td>
<td>–</td>
<td>Ouro Preto do Oeste/RO</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>4</td>
<td>ESJ0H3</td>
<td>C</td>
<td>–</td>
<td>Belém/PA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>5</td>
<td>CP02-1</td>
<td>C</td>
<td>CP-C1</td>
<td>Itajaipe/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>6</td>
<td>Belmont</td>
<td>C</td>
<td>CP-C1</td>
<td>Belmonte/BA</td>
<td>Theobroma sp.</td>
</tr>
<tr>
<td>7</td>
<td>Ilhéus</td>
<td>C</td>
<td>CP-C1</td>
<td>Ilhéus/BA</td>
<td>Theobroma sp.</td>
</tr>
<tr>
<td>8</td>
<td>FA563</td>
<td>C</td>
<td>CP-C1</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>9</td>
<td>Santo Amaro</td>
<td>C</td>
<td>CP-C2</td>
<td>Santo Amaro/BA</td>
<td>Theobroma sp.</td>
</tr>
<tr>
<td>10</td>
<td>FA42</td>
<td>C</td>
<td>CP-C2</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>11</td>
<td>FA276</td>
<td>C</td>
<td>CP-C2</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>12</td>
<td>FA293</td>
<td>C</td>
<td>–</td>
<td>Gandu/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>13</td>
<td>FA562</td>
<td>C</td>
<td>CP-C2</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>14</td>
<td>Lep1</td>
<td>L</td>
<td>–</td>
<td>Pichilingue/Ecuador</td>
<td>Arrabidaeae verrucosa</td>
</tr>
<tr>
<td>15</td>
<td>SCFT</td>
<td>L</td>
<td>–</td>
<td>San Carlos/Ecuador</td>
<td>Arrabidaeae verrucosa</td>
</tr>
<tr>
<td>16</td>
<td>SCL4</td>
<td>L</td>
<td>–</td>
<td>San Carlos/Ecuador</td>
<td>Arrabidaeae verrucosa</td>
</tr>
<tr>
<td>17</td>
<td>FA607</td>
<td>S</td>
<td>–</td>
<td>Coimbra/MG</td>
<td>Solanum lycocarpum</td>
</tr>
<tr>
<td>18</td>
<td>FA609</td>
<td>S</td>
<td>–</td>
<td>Poços de Caldas/MG</td>
<td>Solanum sp.</td>
</tr>
<tr>
<td>19</td>
<td>DOA-105</td>
<td>S</td>
<td>–</td>
<td>Jatari/GO</td>
<td>Solanum lycocarpum</td>
</tr>
<tr>
<td>20</td>
<td>LA17</td>
<td>L</td>
<td>–</td>
<td>Pichilingue/Ecuador</td>
<td>Arrabidaeae verrucosa</td>
</tr>
<tr>
<td>21</td>
<td>RWB500</td>
<td>S</td>
<td>–</td>
<td>Mariana/MG</td>
<td>Solanum cernum</td>
</tr>
<tr>
<td>22</td>
<td>RWB551</td>
<td>S</td>
<td>–</td>
<td>Juiz de Fora/MG</td>
<td>Solanum lycocarpum</td>
</tr>
<tr>
<td>23</td>
<td>FA277</td>
<td>C</td>
<td>–</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>24</td>
<td>FA281</td>
<td>C</td>
<td>CP-C2</td>
<td>Aiquara/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>25</td>
<td>DOA100</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>26</td>
<td>CP02§</td>
<td>C</td>
<td>CP-C1</td>
<td>Itajaipe/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>27</td>
<td>ALF42</td>
<td>C</td>
<td>–</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>28</td>
<td>ALF110</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>29</td>
<td>ALF276</td>
<td>C</td>
<td>–</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>30</td>
<td>ALF277</td>
<td>C</td>
<td>–</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>31</td>
<td>ALF278</td>
<td>C</td>
<td>–</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>32</td>
<td>ALF301</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>33</td>
<td>ALF305</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>34</td>
<td>ALF321</td>
<td>C</td>
<td>–</td>
<td>Ilhéus/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>35</td>
<td>606GD-W</td>
<td>C</td>
<td>–</td>
<td>Itabuna/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>36</td>
<td>676GD-W</td>
<td>C</td>
<td>–</td>
<td>Floresta Azul/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>37</td>
<td>896FD-W</td>
<td>C</td>
<td>–</td>
<td>Jauaquara/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>38</td>
<td>948FD-W</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>39</td>
<td>1734D-W</td>
<td>C</td>
<td>–</td>
<td>Gandu/BA</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>40</td>
<td>FA317</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>41</td>
<td>SABA</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
<tr>
<td>42</td>
<td>1909</td>
<td>C</td>
<td>–</td>
<td>–</td>
<td>Theobroma cacao</td>
</tr>
</tbody>
</table>

*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.
in a final volume of 25 μl containing 1× thermophilic DNA poly Buffer (Promega), 2.5 mM MgCl₂, 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 BanHI, EcoRI, HindIII, KpnI and SalI (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 HindIII or SalI. 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). HindIII-digested DNA was hybridized at 58 °C but that temperature exhibited low specificity for the SalI-digested DNA, making the results difficult to interpret. Subsequently, hybridization with the SalI-digested DNA was performed at 65 °C. Probe labelling, hybridization and detection were performed with the Images Gene 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 Pontecorvo’s minimal medium (Pontecorvo fragments and transferred to 125 ml Erlenmeyer flasks containing 10 days. Once grown, 10 mycelial discs were cut into smaller

**RESULTS**

**Boto transposes in *M. perniciosa***

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 DD²E spacing reported for some transposases of plant *PIF*-like elements, which can also have the DD²E spacing (Zhang *et al.*, 2004). For the *Harbinger* and *Pong* transposases, this motif can be seen as DD²E (Kapitonov & Jurka, 1999, 2004; Zhang *et al.*, 2004).

The first intron (53 bp) in the *Boto* transposase coding region interrupts the His¹¹³ codon and has an A+T content of 68 %. The second intron (48 bp) interrupts the Arg²⁸⁵ codon and has an A+T content of 67 %. The transposase intron positions of *Boto* are different from 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 CPOR1 and CPOR2. 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 EcoRI 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'-GGGCTGGTTCGCTAAAGAAAAGCT- GTAGCCTTTTTCGAGCTTTTC and 5'-GAAAGCTA- CGAAAAAGCTGAGCTTTTTTACCGAACAGGCCC) 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**

Our group had previously designed a phage (2EMBL3) genomic library of *M. perniciosa* aimed at isolating
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 transposases (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: Acrithosiphon pisum (AC202214), Ajellomyces capsulatus (XM_001541700), Aphanomyces euteiches (CU363155), Arabidopsis thaliana (AC005850), Boto (EU218539), Caenorhabditis elegans (NM_062114), Cryptococcus gatti (XM_003102814), Cryptococcus 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).
**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*...
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 SalI enzyme, exhibited from six to 12 copies of the Boto element (Fig. 3d). When analysed with the HindIII 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.
**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. 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 SalI, 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.

**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 Boto2ORF1F1/R1) and 896 or 794 bp (for primers Boto2ORF2F2/R2) using genomic DNA or cDNA, respectively. 'φX', Molecular size marker (DNA from φX174 phage cleaved with HaeIII).
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 II than class I 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 DD48E 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′-GGS-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′-TGTCCGGA-3′) are more similar to PIF (5′-TGGTGGTT-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 & Catcheaside, 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 TCI/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 HindIII DNA fragment is conserved in the M. perniciosa isolates analysed (Fig. 3b, c), a 1.68 kb HindIII 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 (Andebbran 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 TCI-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 retro-transposition 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 (Fundaão de Amparo à Pesquisa do Estado de São Paulo) and SEAGRI (Secretaria de Agricultura do Estado da Bahia) for financial support. Dr. A. Levy helped with the English editing of the manuscript.
REFERENCES


Edited by: R. P. Oliver