Research Paper

**Scatter: a novel family of miniature inverted-repeat transposable elements in the fungus Botrytis cinerea**

Hongyuan Deng1, 2, Dan Shu1, Di Luo1, Tao Gong1, 2, Fenghui Sun1, 2 and Hong Tan1

1 Key Laboratory of Environmental and Applied Microbiology, Chengdu Institute of Biology, The Chinese Academy of Sciences, Chengdu, China
2 University of the Chinese Academy of Sciences, Beijing, China

Miniature inverted-repeat transposable elements (MITEs) are short non-autonomous DNA transposons that play an important role in genome structure and function. Here, we described a novel family of MITEs, named Scatter, identified from the genomes of three strains of the fungus Botrytis cinerea (T4, B05.10, and TBC-A). Intact Scatter elements are typically an average of 247 bp, and contain 41 bp terminal inverted repeats (TIRs) and 2-bp “TA” target site duplications (TSDs).

Based a search against the transposable elements database and GenBank, Scatter is a novel and potentially species-specific family of MITEs. Moderate heterogeneity in sequence and size of individual Scatter copies suggests that Scatter elements were not recently proliferated. Most integrated sites were conserved across all three strains tested and elements inserted at equivalent sites shared high identity at the nucleotide level. This conservation, in combination with the presence of a similar copy number (22–24), in B. cinerea strains tested suggests that Scatter may be a relic of an ancient transposition developed prior to the strain divergence of B. cinerea. Two unique insertion instances were observed, indicating that some copies of Scatter may have remained active following strain divergence of B. cinerea. Because only a few subtle insertion differences among B. cinerea strains were observed, Scatter may play only a minor role in the genetic diversity in B. cinerea species. Most Scatter elements appear to be inserted in potential regulatory regions of adjacent coding regions, highlighting their role in transcriptional regulation. The origin of Scatter remains to be addressed. Scatter is the first well-characterized family of MITEs in B. cinerea.

Additional supporting information may be found in the online version of this article at the publisher’s web-site

**Keywords:** Miniature inverted-repeat transposable elements (MITEs) / DNA transposon / Ancient transposition / Botrytis cinerea

Received: May 3, 2012; accepted: August 25, 2012
DOI 10.1002/jobm.201200238

**Introduction**

Transposable elements (TEs), characterized by their ability to “jump” or transpose within genome, make up a large fraction of the eukaryotic genome and have a substantial impact on genome function and evolution [1–4]. Given the abundance and diversity of TEs, the classification and naming system of TEs has been inconsistent [5]. Although currently used classification systems for TEs largely differ from one another on a lower level, it has been commonly accepted that, at the highest level, TEs can be divided into two classes based on the presence or absence of their transposition intermediate: Class I TEs (or retrotransposons) transpose through reverse transcription of an RNA intermediate, while class II TEs (or DNA transposons) directly “cut and paste” themselves throughout the genome using transposases without an RNA intermediate. Each type of TE contains autonomous and non-autonomous elements. Autonomous elements encode active proteins responsible for their own transposition, whereas non-autonomous elements have no or defective protein coding capacity and are thought to move throughout the genomes by borrowing proteins from autonomous partners [6].

---

Correspondence: Hong Tan, Chengdu Institute of Biology, The Chinese Academy of Sciences, Chengdu 610041, China
E-mail: abath@cib.ac.cn
Phone: +86 28 82890409
Fax: +86 28 82890409

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
www.jbm-journal.com
J. Basic Microbiol. 2013, 53, 815–822
Miniature inverted-repeat transposable elements (MITEs), first discovered in maize [7] and found to be widespread in eukaryotes and prokaryotes [6], are a heterogeneous group of non-autonomous DNA transposons that share structural characteristics, including short length (<600 bp), conserved terminal inverted repeats (TIRs), non-coding capacity of internal sequence, target site duplications (TSDs), high A + T content, tendency to insert into intergenic regions, and the potential to form stable secondary structure [6]. Base on their TIR and TSD sequence signatures, most MITEs can be assigned to one of two major superfamilies, Tourist and Slowaway, although a few ones cannot be classified into a known family [8–10].

Previously, MITEs were most extensively studied in higher eukaryotes and were thought to play an important role in genome evolution and the regulation of gene transcription [1]. Recent reports examining several families of functional small RNA in humans [11], Arabidopsis [12], rice [12], Solanaceae [13], and silkworm [14] derived from MITEs revealed the exceptional role of MITEs in gene regulation. While only a few families were well characterized in fungi, including guest in *Neurospora crassa*, mimp in *Fusarium oxysporum*, and 13 families in *Epichloë festucae* [15–17], studies of these elements provided insights into the potential impact of MITEs on the fungal genome such as genome reshaping attributed to their mobilization or rearrangement mediated by recombination [16, 17] and possible transcriptional regulation of certain gene clusters [17]. Additionally, several uncharacterized MITEs were also recently reported in fungal genomes recently [18, 19], and their further roles remain to be elucidated.

*Botrytis cinerea* (anamorph of *Botryotinia fucheliana*) is a phytopathogenic fungus with significant genetic diversity [20, 21], which is thought to be associated with fungicide resistance and broad host plant fitness [22]. Such genetic diversity may result from TE activity, which is recognized as an important cause of genetic variation, particularly in the organisms without a sexual phase [23]. Previous studies have demonstrated that class I TEs (*Boty*) [24] and class II TEs (*Flipper*) [25] examined in *B. cinerea* partially contributed to genetic variation between the *B. cinerea* populations tested [26]. Although MITEs were also discovered in *B. cinerea* strains [27], their structure, classification, or impact on genome evolution and gene expression have not been characterized.

Here we described a family of MITEs termed Scatter in *B. cinerea* genomes. Little similarity with any known MITEs indicates that Scatter is a new MITE family. The distribution in the *B. cinerea* genome, the possible impact on genome evolution and function, as well as the origin of this novel family of MITEs was further analyzed. Scatter is the first well-characterized family of MITEs in *B. cinerea*. Materials and methods

**Genome database of *B. cinerea* strains T4, B05.10, and TBC-A for in silico analysis**

The 4.5× sequencing coverage whole genome shotgun (WGS) sequences of *B. cinerea* strain B05.10, which was obtained after benomyl treatment of a *Vitis* isolate [28], are publicly available (http://www.broad institute.org/ annotation/genome/botrytis_cinerea.2/Home.html). The genome of *B. cinerea* strain T4, which was isolated from tomato plants grown in a glasshouse [29], was sequenced using 10× coverage, and the genome data was deposited to GenBank (GenBank ID: 64593) and the URGI website (http://urgi.versailles.inra.fr/Species/Botrytis/Sequences-Databases). The genome of *B. cinerea* strain TBC-A, which was originally isolated from wheat in China and is of great industrial interest, was sequenced by our lab using 20× coverage (data not published).

**Mining and characterization of Scatter elements from *B. cinerea* genomes**

Two TEs-like sequences located at the 5′ upstream sequence of a gene encoding putative sesquiterpene cyclase (accession no. Bofu T4_P062370.1) were used to search against the *B. cinerea* genome database using BLASTN. All hits with an E-value < e−6 were extracted for further analysis. Intact elements with 10 bp stretches flanking both ends were multiple aligned using Clustal X [30]. A consensus sequence was obtained by calculating the frequencies of nucleotide base usage at respective positions of given sequences as follows: a base showing a frequency more than 50% was considered as a consensus base; if not, all abundant bases showing a total of more than 50% frequency were denoted as consensus. To guarantee that all Scatter were obtained, the consensus sequence was used to search the genome database again using an E-value cutoff e−6.

Blt2seq (www.ncbi.nlm.nih.gov/Blast/bl2seq/wBlast2.cgi) was performed to identify TIRs. The ORFfinder online tool (http://www.ncbi.nlm.nih.gov/orffinder.html) was used to analyze coding capacity. DNA secondary structures were predicted by SantaLucia free energy minimization algorithms [31] using the web Mfold server with default settings [32].

To determine whether Scatter elements are related to any known TEs, the ISfinder database (http://www-is.biotoul.fr/) [33] and Repbase Update (version 17.02) [34] were searched. The NCBI nr database was searched to identify homologous copies in other species.

The TIR sequence was extracted to search against the NCBI nr database using BLAST with E-value threshold e−6, and subsequently the surrounding sequence of the...
hits were subjected to BLASTX analysis to determine the presence of putative autonomous parent or partner elements.

The potential transcription activity of Scatter was examined by searching the B. cinerea EST database at NCBI using BLASTN.

**Distribution and integrated site tendency analysis of Scatter**

Each Scatter element was mapped to supercontig or scaffold to identify their location in the genome. The location of Scatter and the distance between Scatter elements and the most closely adjacent open reading frames (ORFs) were identified by manual inspection. A comparison of Scatter integration sites among B. cinerea strain T4, B05.10, and TBC-A were also carried out as follows: each Scatter element with 200-bp flanking fragments at both ends was extracted and then searched against two genome database. The sequence identity of matched pairs of elements sharing the identical insertion site was investigated by pairwise alignment using the Bt2seq tool (www.ncbi.nlm.nih.gov/Blast/bl2seq/wBlast2.cgi).

**Results**

**B. cinerea genome mining for TE-like elements: Scatter**

The first two Scatter elements were revealed from genome of B. cinerea T4 strain during analysis of 5’ upstream sequence of a gene encoding putative sesquiterpene cyclase (accession no. Bofu T4_P062370.1). Two Scatter sequences (designated as Scatter-T2 and Scatter-T3), located 600 bp upstream of the start codon, share high sequence similarity (84%) and are separated by a 71-bp fragment. The presence of nearly perfect TIRs and flanking 2-bp direct repeats (Fig. 1) suggested that both of these two elements may be TEs. Next, an NCBI BLASTN search against the B. cinerea T4 genome database was performed using the two possible TEs as the query, and 23 hits were identified with E-values < e^-6. Each hit was classified into two groups; sequences with both complete TIRs was regarded as intact and otherwise regarded as truncated. Seventeen hits were intact and 6 hits were truncated. Using a similar approach, 24 (15 intact and 9 truncated homologous sequences, designated as Scatter-B1 to Scatter-B24) and 22 (17 intact and 5 truncated, designated as Scatter-G1 to Scatter-C22) were obtained from B. cinerea strain B05.10 and TBC-A, respectively. Intact elements (Supporting Information Tables S1 and S2) from three B. cinerea strains were further analyzed to obtain structure details.

**Characterization of Scatter elements in B. cinerea**

In the B. cinerea T4 strain, intact Scatter elements are 229–260 bp (247 bp on average) in length. These sequence are rich in AT content (range from 57.6 to 69.8%, 64.0% on average), which is significantly higher than the average A + T base composition (56.8%) of the whole genome. Multiple sequences alignments showed that these elements share the 41-bp TIR and are flanked by a clearly recognizable 2-bp “TA” TSDs (Fig. 1). No ORF could be identified in the region bound by the two TIR signals of each copy, indicating that they are non-autonomous TEs. Insertion site analysis showed that these sequences preferentially insert into the intergenic region. DNA folding predictions suggested that all of these elements have the potential to form secondary structures. The stability of the predicted structures, expressed as thermodynamic values (ΔG), range from −48.28 kcal mol⁻¹ (Scatter-T8) to −22.07 kcal mol⁻¹ (Scatter-T1), with an average of −37.2 kcal mol⁻¹.

In both B. cinerea B05.10 strain and TBC-A strain, the structural characteristics (Table 1) of intact elements are organized roughly in the same manner as in the T4 strain, including short length, conserved TIRs and TSDs, non-coding capacity of internal sequence, high A + T content, potential to form secondary structure, and a bias to insert into the intergenic region. These features of intact elements described above correspond with those of known MITEs previously identified in other species [6], suggesting that Scatter elements are typical MITEs.

Multiple sequence alignment of intact Scatter elements for each B. cinerea strain tested also revealed moderate sequence variations and some indels (Fig. 1) between individual copies, implying that Scatter elements were not recently proliferated since newly replicated elements would show high homogeneity in sequence and size [17].

Because MITEs are non-autonomous elements and their transposition requires involvement of autonomous elements, the surrounding region of sequences sharing high similarity with the TIR in each strain tested were analyzed to identify the putative parent or partner elements. However, none showed significant similarity with any known or putative DNA transposase responsible for transposition of DNA transposon. This may suggest that Scatter is no longer active.

To categorize Scatter elements, a search against the TEs database based on the consensus sequence of Scatter was carried out, and the results showed that Scatter does not exhibit significant sequence identity with any other known TEs. Further detailed comparisons revealed that
Figure 1. Multiple alignments of intact Scatter elements in B. cinerea T4 strain. TIR is the terminal inverted repeat indicated by solid arrow. TSD is the target site duplication indicated by hollow arrow. Consensus sequence was obtained according to the method as described in Materials and Methods Section and $M = A/C, Y = T/C, W = A/T$. 

it shares the same TA di-nucleotides TSDs with three well-characterized MITEs, including mimp, Flipper, and Stowaway; the former two elements were discovered in fungi [16, 25] and the latter was recognized as one of the two predominant MITEs families in plants [35]. However, Scatter is significantly different from those elements in TIR, which was considered as recognition signal of transposase and a key classification signature of the MITEs family [6]. Therefore, Scatter is a novel MITEs family in B. cinerea.

Additionally, based on the results from the GenBank nr database search, Scatter elements were only identified in B. cinerea strains tested, and were not present in any other species, even in the closely related fungi Sclerotinia sclerotiorum. These data suggested that Scatter may be a species-specific MITE family of B. cinerea.

MITEs can regulate gene expression through small RNA pathways in higher eukaryotes as reported previously [11–14]. To gain insight regarding whether Scatter elements act in such a way, we investigated their transcription potential by searching against the EST library of B. cinerea T4. No hit with significant similarity was obtained, suggesting that Scatter may not involve in small RNA pathways.

### Integrated sites analysis of Scatter elements

To examine possible impacts of Scatter elements on the genome, precise integrated sites of Scatter elements were analyzed and the exact distance between each element and the nearest ORF was calculated. As shown in Fig. 2B, approximately 60% of integrations were found within 500 bp upstream or downstream to ORFs, indicating that they may have a regulatory function on gene transcription. This integrated tendency highlighted its potential effect on adjacent gene expression. Additionally, because MITEs in the fungus E. festucae were found to be non-randomly enriched in nearby secondary metabolite genes [17], each predicted protein of nearest ORF of Scatter elements was examined using BLASTX search. The results showed that all predicted proteins close to Scatter elements do not exhibit any obvious relevance or common attributes, thus ruling out this tendency.

Further, integrated sites of all Scatter elements including intact and truncated sites were compared between all three B. cinerea strains tested. Twenty-one integrated sites (accounted for 91, 87, and 95% of the total number of Scatter in strain T4, B05.10, and TBC-A, respectively) were found to be conserved across the genomes tested (Fig. 2A), and the sequences of the elements inserted in equivalent positions share identity as high as 98.7%, although these

### Table 1. Characteristics of Scatter in three B. cinerea genomes.

<table>
<thead>
<tr>
<th>B. cinerea strain</th>
<th>Copy number (intact/truncated)</th>
<th>Mean length of intact Scatter (range) bp</th>
<th>Mean A + T (range)%</th>
<th>TIR length (bp)</th>
<th>Mean ΔG (kcal mol⁻¹) ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4</td>
<td>23 (17/6)</td>
<td>247 bp (229–260 bp)</td>
<td>64.0% (57.6–69.8%)</td>
<td>TA</td>
<td>41</td>
</tr>
<tr>
<td>B05.10</td>
<td>24 (15/9)</td>
<td>246 bp (229–260 bp)</td>
<td>65.7% (56.7–72.3%)</td>
<td>TA</td>
<td>41</td>
</tr>
<tr>
<td>TBC-A</td>
<td>22 (17/5)</td>
<td>247 bp (229–260 bp)</td>
<td>64.1% (57.3–69.8%)</td>
<td>TA</td>
<td>41</td>
</tr>
</tbody>
</table>

¹ ΔG represents thermodynamic value indicating the stability of secondary structure of Scatter element.
strains were isolated from different regions and host plants [28, 29]. This conservation suggests that these elements may have transposed prior to strain divergence of *B. cinerea*.

Interestingly, two unique insertions were observed upon comparative analysis of the integrated sites. While Scatter-B14 inserted into the supercontig_1.103 of the B05.10 genome, Scatter-C17 integrated into the scaffold_12 of the TBC-A genome. Each element is absent in the corresponding loci of the two other genomes, although the sequence flanking the integrated site matches perfectly in all three genomes tested. Thus, a portion of elements remained active after strain divergence of *B. cinerea*.

**Discussion**

In this study, we first described a family of MITEs termed Scatter from the fungus *B. cinerea* using *in silico* analysis. Based on sequence characteristics and comparative analysis with other known MITEs, Scatter appears to be a novel family of MITEs. This family has only been identified in *B. cinerea*, indicating that it may be a species-specific family of MITEs and has the potential to be applied as a molecular marker for identifying *B. cinerea* species.

Moderate heterogeneity in sequence and size of individual Scatter copies suggests Scatter elements were not recently proliferated. The presence of a portion of truncated elements in each strain tested also supports this postulation. Sequence variation was unlikely to have resulted from repeat induced point-mutation (RIP) because all members of the Scatter family are smaller than the 400 bp identified in *N. crassa* as the minimum length for RIP to function [36] and no obvious bias toward C:T or G:A transitions characteristic of RIP was observed [36]. Thus, it was initially presumed that these sequence variations may have resulted from basal levels of mutations over a long period of time. However, most integrated sites were conserved across all three *B. cinerea* strains tested, and the elements inserted at equivalent sites share high identity at the nucleotide level. These data, along with the presence of similar copy number in each genome, collectively suggest that Scatter elements may be relics of ancient transposition events prior to strain divergence of *B. cinerea*. Thus, most sequence variations and indels between copies in each genome existed long before divergence with only a small portion attributed to basal mutation.

Two unique insertion instances that were not present in a second *B. cinerea* genome implied that a portion of Scatter elements may be still active after strain divergence. However, further investigation revealed no additional differential reinsertions or excisions “footprints,” which may reflect the capacity of mobilization of MITEs, between these three genomes tested. Based on the absence of parent or partner autonomous elements that drive the mobilization of Scatter, the high conservation of integrated sites, and equivalent sequences, these sequences appear to have been inactivated soon after transposition.

Generally, MITEs can arise in two ways. The first and major way is from internal deletion of full-length autonomous elements [37], with MITEs originating in such a way that result in exhibit extended regions of similarity with parental elements. This is the case for several well-characterized MITEs from high eukaryotes [6] and for several MITEs in fungi [17, 38], such as guest, which were identified as deleted derivates of Tc1/mariner elements in *N. crassa* [38]. The second method is to evolve independently following association of TIRs bordering a segment of genomic DNA [39], from which MITEs would show limited similarity with TIRs. And thus far only a few elements, such as a Ds1 in grass Poaceae [40], were found to be created through this *de novo* pathway. Multiple alignments between intact Scatter elements showed extensive similarity in both TIRs and internal regions, indicating that Scatter may have originated from the first way. However, no possible parental autonomous element of Scatter, whose existence is direct evidence of the first origin of Scatter, was found, as all intact Scatter elements were non-coding and the hypothetical protein encoded by sequence adjacent to TIR and TIR-like sequence in each strain tested showed little similarity with any known or predicted transposase. Thus, the origin of Scatter remains to be addressed to identify all possible autonomous DNA transposons in *B. cinerea* and evaluate their binding capacity with TIRs of Scatter. Screening newly uncovered genomes to identify new Scatter or Scatter-like elements and related autonomous elements should also be conducted.

TEs are considered an important cause of genetic variation due to their mobilization and further chromosome rearrangement mediated by recombination [3]. Previous studies demonstrated that two autonomous elements, *Flipper* and *Boty*, contribute to the genetic diversity of *B. cinerea* species [24, 25]. Our analysis showed that Scatter in *B. cinerea* did not lead to dramatic alterations in genome structure, and that only a few subtle insertion differences among *B. cinerea* strains were observed. Furthermore, most Scatter elements appeared to have inactivated after strain divergence of *B. cinerea*. Therefore, Scatter may play only a minor role in the
genetic diversity of \textit{B. cinerea} species compared to \textit{Boty} and \textit{Flipper} \cite{24, 25}.

Approximately 60\% integrations were found within 500 bp upstream or downstream to ORFs, which are likely to be regulatory regions of gene transcription. As transposons in other organisms have been shown to modulate gene expression by disrupting existing or generating new cis-elements, or through altering the chromatin environment \cite{1}, this integrated tendency of \textit{Scatter} highlighted its potential effect on adjacent gene expression. Whether and how \textit{Scatter} affects gene transcription in \textit{B. cinerea} cannot be determined from direct analysis of genome database but should be examined through comparative transcription analysis between \textit{B. cinerea} wild type strains and \textit{Scatter} deletion mutants using knock-out or other approaches.

**Conflict of Interest**

All authors have no conflict of interest to declare.

**Acknowledgments**

This work was funded by grants from the Knowledge Innovation Program of the Chinese Academy of Sciences (No. KSCX2-EW-J-22 and KSCX2-EW-G-16-01), the Frontier Program of Chengdu Institute of Biology, the Chinese Academy of Sciences, and West Light Foundation of the Chinese Academy of Science (No. Y1C5031).

**References**

research of transposable elements in filamentous fungi.


