Evolutionary History of the impala Transposon in Fusarium oxysporum

Aurélie Hua-Van, Thierry Langin, and Marie-Josée Daboussi
Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay, France

impala is an active DNA transposon family that was first identified in a strain of Fusarium oxysporum pathogenic to melon. The 10 copies present in this strain define three subfamilies that differ by about 20% at the nucleotide level. This high level of polymorphism suggests the existence of an ancestral polymorphism associated with vertical transmission and/or the introduction of some subfamilies by horizontal transfer from another species. To gain insights into the molecular evolution of this family, impala distribution was investigated in strains with various host specificities by Southern blot, PCR, and sequencing. Detection of impala elements in most of the F. oxysporum strains tested indicates that impala is an ancient component of the F. oxysporum genome. Subfamily-specific amplifications and sequence and phylogenetic analyses revealed five subfamilies, several of which can be found within the same genome. This supports the hypothesis of an ancestral polymorphism followed by vertical transmission and independent evolution in the host-specific forms. Highly similar elements showing unique features (internal deletions, high rates of CG-to-TA transitions) or being present at the same genomic location were identified in several strains with different host specificities, raising questions about the phylogenetic relationships of these strains. A phylogenetic analysis performed by sequencing a portion of the EF1α gene showed in most cases a correlation between the presence of a particular element and a close genetic relationship. All of these data provide important information on the evolutionary origin of this element and reveal its potential as a valuable tool for tracing populations.

Introduction

Fusarium oxysporum is a complex of soil-borne fungi responsible for diseases in more than 120 plant species. This asexual species complex is composed of numerous strains classed into several formae speciales based on pathogenic criteria (Armstrong and Armstrong 1981). Each forma specialis groups strains that are pathogenic to one particular species or group of plants. Formae speciales are further subdivided into physiological races according to cultivar specificity. Phylogenetic relationships in F. oxysporum have been the subject of many studies undertaken to investigate evolutionary trends among species and to determine the dynamics of the different formae speciales and physiological races (Gordon and Martyn 1997). Genetic distances among strains have been evaluated through analyses of pathogenicity, vegetative compatibility group (VCG), chromosomal features, rDNA restriction fragment length polymorphism, mtDNA, and other molecular markers (Jacobson and Gordon 1990; Appel and Gordon 1995, 1996; O’Donnell et al. 1998; Alves-Santos et al. 1999). Most studies reported to date have focused on an individual forma specialis, seeking to characterize the diversity therein, especially as it relates to physiological races. In many cases, it was found that formae speciales are genetically heterogeneous, and can sometimes have a polyphyletic origin (O’Donnell et al. 1998).

More recent studies have used new tools such as fingerprinting with repeated sequences (usually transposable elements), proven to be valuable for several fungal species, including Magnaporthe grisea (Hamer et al. 1989; Dobinson, Harris, and Hamer 1993; Farman, Taura, and Leong 1996), Mycosphaerella graminicola (McDonald and Martinez 1990), Erysiphe graminis (O’Dell et al. 1989), Cryphonectria parasitica (Milgroom, Lipari, and Powell 1992), and F. oxysporum. In this species, fingerprinting allowed researchers to distinguish formae speciales (Namiki et al. 1994), to track the origin of new infestation (Mouyna, Renard, and Brygoo 1996), to detect a given forma specialis (Fernandez et al. 1998), and to identify races within a forma specialis (Chioicchetti et al. 1999). Moreover, PCR assays based on several transposable element insertion sites provided a useful diagnostic tool for quickly identifying formae speciales and races.

Transposable elements appear especially abundant within F. oxysporum that exhibit a high degree of genetic variability. This is illustrated by numerous transposon families (about 17) characterized thus far, representing the major classes of retroelements and DNA transposons (Julien, Poirier-Hamon, and Brygoo 1992; Daboussi and Langin 1994; Mouyna, Renard, and Brygoo 1996; Okuda et al. 1998; Gomez-Gomez et al. 1999; Hua-Van et al. 2000; Mes, Haring, and Cornelissen 2000). One of these active DNA transposon families, named impala, is composed of few (8–10) elements and is typically 1,280 bp long with 37-bp inverted terminal repeats (ITRs). Impala contains a single open reading frame encoding a transposase of 340 amino acids (aa). This transposase is related to those found in elements belonging to the widespread Tc1-mariner superfamily (Langin, Capy, and Daboussi 1995; Hua-Van et al. 1998). In strain FOM24 (herein called M24), in which impala was first identified, about 10 different impala copies have been characterized. Three subfamilies, named E, D, and F, have been detected (Hua-Van et al. 1998). The E and D subfamilies are represented by several copies, which are autonomous, inactive, or truncated (Hua-Van et al. 1998; Hua-Van et al. 2001). Within each subfamily, nucleotide divergence between the full-

1 Present address: Institut de Biotechnologie des Plantes, Université Paris-Sud, Orsay, France.

Key words: transposable element, filamentous fungi, Fusarium oxysporum, Tc1-mariner superfamily, evolution, repeat-induced point mutation.

Address for correspondence and reprints: Marie-Josée Daboussi, Institut de Génétique et Microbiologie, Université Paris-Sud, 91405 Orsay cedex, France. E-mail: daboussi@igmors.u-psud.fr.

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
length truncated elements is relatively low (around 1%), while the truncated elements are more polymorphic. The F subfamily contains only one full-length-but-inactivated element. These three subfamilies differ by as much as 20% at the nucleotide level (Hua-Van et al. 1998). This result is intriguing when compared with the 0.3%–5% polymorphism observed within the *F. oxysporum* complex for the *nia* and EF1α genes and internal transcribed sequence or intergenic sequence of ribosomal DNA (Avelange 1994; Appel and Gordon 1996; O’Donnell et al. 1998). Two non-mutually-exclusive hypotheses may be proposed to explain the evolutionary origin of impala subfamilies. First, the presence of different subfamilies in a genome might be the result of ancestral polymorphism. The subfamilies present in the common ancestor would then be expected to be present in genetically diverse strains associated with diversification of these pathogens. Another possibility is the occurrence of one or more horizontal transfers. Such events have been described for other transposable elements (Kidwell 1992; Robertson and Lampe 1995), notably for mariner elements, for which horizontal transfer appears to have played a major role in evolution (Garcia-Fernandez et al. 1995; Lohe et al. 1995; Robertson and Lampe 1995b). In this case, the foreign impala element(s) might be expected to be present in a small number of related strains, all derived from the same ancestor in which the transfer occurred.

In order to understand the dynamics of impala within the complex, a collection of *F. oxysporum* strains with different host specificities were analyzed by Southern blot, PCR, and sequencing. This analysis showed that the impala family is widespread within *F. oxysporum* strains. A total of five subfamilies were identified, and most of these were present in several strains with different host specificities. On several occasions, similar inactive copies with unique sequence characteristics were detected in strains with various host specificities. Relationships of these strains were inferred through phylogenetic analysis of translation elongation factor (EF1α) gene sequences. This revealed that strains containing very similar impala copies are usually closely related. The implications of these results for the evolutionary origin of strains and the use of impala as a tool are discussed.

**Materials and Methods**

**Fungal Strains**

Fungal strains used in this study are listed in table 1. Strains were obtained from the following sources: (1) C. Alabouvette, Institut National de Recherche Agronomique, Dijon, France; (2) T. R. Gordon, Department of Plant Pathology, University of California, Davis, Calif.; (3) D. Fernandez, Institut de Recherche pour le Développement, Montpellier, France; (4) J. Guadet, Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay, France; and (5) R. C. Ploetz, Tropical Research and Education Center, University of Florida, Homestead, Fla.

DNA Extraction and Southern Blot Analysis

DNA extractions were conducted as previously described in Langin et al. (1990). Ten micrograms of genomic DNA was digested with *Eco*RI, separated by 0.8% agarose gel electrophoresis, transferred on nylon membrane (Amersham) using a vacuum blotter, and hybridized using probes labeled with the Pharmacia T7 Quick Prime Kit according to standard procedures (Sambrook, Fritsch, and Maniatis 1989).

**Polymerase Chain Reaction and Primer Sequences**

Three pairs of primers were used to specifically amplify elements of the E subfamily (SpE5 [AGAACCACAATGCCGCGG] and SpE3 [TCCGGCCCA-TATGCACAGAG]) or the D subfamily (SpD5 [AGGGGTTACCGCACTCACAG] and SpD3 [CAGGGCCA-GTGGAAACAGAC]) or any impala elements (NS5 [CGATGCTCTCAAGGCGAAGGAA] and NS3 [TCTGCTCCTCATCAACGCG]). Elements at particular locations were amplified using primers nested in the flanking regions, about 100 bp from either side of the elements: SFA5 (TGTGCCCACTTTTGTTCTG) and SFA3 (TATGGCAGCTTAGAAAGTCCCG) for impA. SFC5 (ACCACTTTTAGATGCTCGG) and SFC3 (ATCTAAACAAGGGGTGCGCG) for impC, and SFD5 (ATCTTGTTGTATTCTCTGACCC) and SFD3 (ACACCCGATTTCCCCACTACG) for impD. For impB and impG amplifications, a primer in the flanking region and an internal primer were used: SFB5 (CAATTCATTGCA-CCGTTTCCCTCACGC) and SpE3 for impb, or SFG3 (TCTGAG-GAAAGAATCTGATC) and SpG5 (TCTGTTGAG-ATGTTGGAGGG) for impG. The latter primer primes on the solo-LTR Han, inserted in impG (Hua-Van et al. 2000). Amplification of the EF1α fragment was conducted using primers EF-1 and EF-2 described in O’Donnell et al. (1998). Amplifications were done using standard procedures in a PTC100 thermocycler (MJ Research, Waldham, Mass.) using the following program: 30 cycles of 1 min at 94°C, 30 s at 60°C, and 1 min 30 s at 72°C, followed by a final extension of 10 min at 72°C.

**Inverse PCR**

To determine if deleted copies or ripped copies originated from the same event, the insertion points of these copies were cloned by Inverse PCR in strain MK, following a procedure adapted from Gloor et al. (1993). One hundred nanograms of genomic DNA was digested with restriction enzyme for 30 min at room temperature in a total volume of 20 μl. After 10 min inactivation at 65°C, 3 μl was used in a ligation experiment (10 μl final volume with Promega ligase and buffer; 30 min at room temperature). The ligation product inactivated for 10 min at 65°C was used in a PCR experiment using the following PCR program: 2 cycles of 95°C for 1 min, 58°C for 15 s, and 72°C for 5 min; 35 cycles of 95°C for 15 s, 58°C for 15 s, and 72°C for 3 min; and 72°C for 10 min.

For deleted copies, we used the restriction enzyme *Xba*I and primers DDV5 (AABACCGTGCTACGCT-
Table 1
Strains Used in this Study and impala Distribution

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>ABB.</th>
<th>F. SP.</th>
<th>HOST</th>
<th>ORIGIN</th>
<th>SB</th>
<th>NS</th>
<th>SpE</th>
<th>SpD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOM24</td>
<td>M24</td>
<td>melonis</td>
<td>Melon</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FOM150</td>
<td>M150</td>
<td>melonis</td>
<td>Melon</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FOM7</td>
<td>M7</td>
<td>melonis</td>
<td>Melon</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FOMK419</td>
<td>MK</td>
<td>melonis</td>
<td>Melon</td>
<td>Mexico</td>
<td>(2)</td>
<td>+</td>
<td>+*</td>
<td>+</td>
</tr>
<tr>
<td>FOMP2</td>
<td>MP</td>
<td>melonis</td>
<td>Melon</td>
<td>California</td>
<td>(2)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOM1127</td>
<td>M11</td>
<td>melonis</td>
<td>Melon</td>
<td>New York</td>
<td>(2)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FO1</td>
<td>S5</td>
<td>NP</td>
<td>Soil</td>
<td>France</td>
<td>(1)</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>FO47</td>
<td>S47</td>
<td>NP</td>
<td>Soil</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOLn3</td>
<td>Ln3</td>
<td>lini</td>
<td>Flax</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FOLn68248</td>
<td>Ln86</td>
<td>lini</td>
<td>Flax</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOLn88321</td>
<td>Ln88</td>
<td>lini</td>
<td>Flax</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FOL15</td>
<td>L15</td>
<td>lycopersici</td>
<td>Tomato</td>
<td>Tunisia</td>
<td>(1)</td>
<td>+</td>
<td>+*</td>
<td>−</td>
</tr>
<tr>
<td>FORL28</td>
<td>RL28</td>
<td>rad-lyco.</td>
<td>Tomato</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+*</td>
<td>+</td>
</tr>
<tr>
<td>FORL4</td>
<td>RL4</td>
<td>rad.-lyco.</td>
<td>Tomato</td>
<td>Canada</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOA1</td>
<td>A1</td>
<td>albedinis</td>
<td>Date palm</td>
<td>Morocco</td>
<td>(3)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOA4</td>
<td>A4</td>
<td>albedinis</td>
<td>Date palm</td>
<td>Morocco</td>
<td>(3)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOA5</td>
<td>A5</td>
<td>albedinis</td>
<td>Date palm</td>
<td>Morocco</td>
<td>(3)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOA438</td>
<td>A438</td>
<td>albedinis</td>
<td>Date palm</td>
<td>Morocco</td>
<td>(3)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOB</td>
<td>B</td>
<td>bulbigenum</td>
<td>Vanilla</td>
<td>(4)</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>FOca cal</td>
<td>Cac</td>
<td>canariensis</td>
<td>Orn. palm</td>
<td>California</td>
<td>(1)</td>
<td>−</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOca o1</td>
<td>CaO</td>
<td>canariensis</td>
<td>Orn. palm</td>
<td>France</td>
<td>(1)</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>FOca lago</td>
<td>Cai</td>
<td>canariensis</td>
<td>Orn. palm</td>
<td>Canary</td>
<td>(1)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOci</td>
<td>Ci</td>
<td>ciceri</td>
<td>Chickpea</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>+***</td>
<td></td>
</tr>
<tr>
<td>FOcu2</td>
<td>Cu</td>
<td>cubense</td>
<td>Banana tree</td>
<td>Costa Rica</td>
<td>(5)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOD11</td>
<td>D11</td>
<td>dianthi</td>
<td>Carnation</td>
<td>Italy</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOg</td>
<td>G</td>
<td>gladolius</td>
<td>Gladiolus</td>
<td>(4)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>FOP108</td>
<td>P108</td>
<td>phaseoli</td>
<td>Com. bean</td>
<td>Brazil</td>
<td>(4)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FOP204</td>
<td>P204</td>
<td>phaseoli</td>
<td>Com. bean</td>
<td>Columbia</td>
<td>(4)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FO var red</td>
<td>redolens</td>
<td>radolens</td>
<td>Radish</td>
<td>France</td>
<td>(1)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FOR4</td>
<td>R4</td>
<td>raphani</td>
<td>Radish</td>
<td>Germany</td>
<td>(1)</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>FOR5</td>
<td>R5</td>
<td>raphani</td>
<td>Radish</td>
<td>USA</td>
<td>(3)</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>FORv1</td>
<td>V</td>
<td>vastifunctum</td>
<td>Cotton</td>
<td>USA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.—F. sp. = formae speciales; SB = Southern blot; NP = nonpathogenic; ND = not determined; orn. = ornamental; com. = common; rad.-lyco. = radicis lycopersici; + = detection of hybridization signal or PCR amplification; − = absence of signal or amplification; * = fragment shorter than expected; ** = fragment larger than expected.

a The abbreviated names of the strains (used throughout the text).

b Strains were provided by different laboratories (numbers in parentheses). (1) C. Alabouvette, Institut National de Recherche Agronomique, Dijon, France; (2) T. R. Gordon, Department of Plant Pathology, University of California, Davis, Calif.; (3) D. Fernandez, Institut de Recherche pour le Développement, Montpellier, France; (4) J. Guadet, Institut de Généétique et Microbiologie, Université Paris-Sud, Orsay, France; (5) R. C. Ploetz, Tropical Research and Education Center, University of Florida, Homestead, Fla.

c Primer pairs used in PCR experiments correspond to “nonspecific” NS3 and NS5 primers (NS), allowing amplification of any SpE3 and SpE5, specific for subfamily E (SpE), and primers SpD3 and SpD5, specific for subfamily D (SpD). These primers are described in Materials and Methods.

d redolens strains are considered as a species distinct from Fusarium oxysporum (see O’Donnell and Cigelnik 1997).

Cloning and Sequencing

Impala PCR products were cloned using the Pro-mega pGEMT-easy kit. For each strain, several individual clones were sequenced using the ABI PRISM DyeTerminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) on an ABI373 Automated DNA sequencer (Applied Biosystems). Sequencing primers corresponded to M13 forward and M13 reverse primers. EF1α sequencing was performed either after cloning as described above or directly on PCR products using primers EF-1 and EF-2. The nucleotide sequences reported in this paper are deposited in GenBank under accession numbers AF363407–AF363439 (impala) and AF363394–AF363406 (EF1α).

Phylogenetic Analysis

Multiple DNA alignments were obtained with the Clustal V program (Higgins and Sharp 1988). Maximum-parsimony analyses were performed using the heuristic algorithm and default options of PAUP 4.0
We carried out 100 or 1,000 bootstrap replications of maximum parsimony with a full heuristic search and default options.

**Results**

**Copy Number and Genomic Distribution of impala**

Southern blot experiments using a copy of imp160 or a mix of PCR products (impEDF; see below) as probes were performed on 32 strains belonging to 14 *formeae speciales* (table 1 and fig. 1). These experiments revealed that all but three strains contained impala elements. The number of hybridization signals (≥10) indicated that impala was always present in low copy numbers. The variation in intensity of the hybridization signals (fig. 1A) probably reflects the presence of several divergent subfamilies in the same genome, as demonstrated in strain M24 (Hua-Van et al. 1998). Similar variations (fig. 1B) were obtained with the use of the mixed probe impEDF, composed of 840-bp impala internal PCR fragments obtained from copies of impE, impD, and impF, which are representative of the three known subfamilies (see below), indicating that other subfamilies may exist.

PCR experiments were carried out on a set of 23 strains, representing 13 *formeae speciales*, using a set of specific primers for the major impala subfamilies. Nonspecific (NS) primers able to amplify all types of impala were used as a positive controls (fig. 2A). As shown in figure 2B and table 1, all strains showing impala sequences by Southern blot gave PCR products using NS primers. Using the specific primers (SpE and SpD pairs), we were able to detect elements belonging to E and to D subfamilies in 17 and 10 strains, respectively. Eight strains contained both subfamilies, and one strain contained none of these subfamilies, although impala elements were detected with NS primers. Several strains gave additional PCR products of unexpected sizes. Shorter fragments were obtained in *formeae speciales* melonis (MK), lycopersici (L15), and radicis-lycopersici (RL28) using the NS primers and the SpD primers which are specific for the D subfamily, strongly suggesting the presence of internally deleted elements of the D subfamily in these strains. In *formeae speciales lini* (Ln3) and ciceri (Ci), a longer fragment was obtained with the SpD primers but not with the NS primers. This could reflect the existence of more complex rearrangements within the impala elements.

When the PCR products were hybridized with the probe impEDF (see above and fig. 2A and B), strong hybridization signals were observed in some strains (i.e., M24 and MK), while other strains (i.e., MP and RL4) contained elements that gave very weak hybridization signals. These elements may represent new subfamilies or highly polymorphic members of a known subfamily.

**Sequence Variation and Phylogenetic Analysis**

The different PCR products obtained with the NS primers were cloned, and some of these clones were completely sequenced. A phylogenetic tree deduced from these sequences by maximum parsimony is shown in figure 3. The clades corresponding to the known subfamilies E, D, and F were strongly supported by bootstrapping (Hua-Van et al. 1998). The E subfamily (100% bootstrap) is composed of several groups of highly similar elements (around 4% nucleotide divergence; divergence can reach 12% for more divergent members), except for elements RL4-110, M11-6, and impB, which are supported by long branches. The D subfamily (100% bootstrap) is also very homogeneous in sequence (2.3% sequence divergence) but is characterized by the presence of several internally deleted elements. The F subfamily (98% bootstrap), for which only one element from strain M24 had been described, was also found in two other strains and appeared less homogeneous (14% nucleotide divergence).
The phylogenetic analysis resulted in the identification of two additional subfamilies: one of them, called the K subfamily (63% bootstrap), is composed of five sequences obtained from four different formae speciales. The two subclades within the K subfamily received 100% bootstrap support, and one of these, composed of three sequences, was supported by the longest node within the tree. These three elements presented a particular mutation pattern (see below) that can explain this high nucleotide divergence (an average of 14% between the two subclades). Nevertheless, these copies appeared closer to the two other sequences within the K subfamily and have been considered to belong to this subfamily.

The fifth subfamily, P, contained only one member identified in strain MP. The subfamilies differ from each other by 20%–30% nucleotide divergence, with the P subfamily being the most divergent (26%–30%).

Active Versus Inactive Elements

Potentially active elements (i.e., containing no stop codons or frameshift mutations over the 800 bp sequenced) were found within the E and D subfamilies (fig. 3), similar to those described previously from strain M24 (Hua-Van et al. 1998). The present study also revealed the existence of such elements in the F subfamily but not in the K and P subfamilies.
Most of the sequences represented inactive copies, as deduced from the presence of stop codons and/or frameshifts. They were found in every subfamily. Among these, we identified two particular types of inactive elements. The first type was characterized by internal deletions. Three sequences with deletions, RL28\(\Delta\)22, L15\(\Delta\)5, and MK\(\Delta\)208 (belonging to the D subfamily), were identified in three different *formae specialae* (fig. 3). Surprisingly, each of these possessed the same 327-bp deletion, and these truncated elements were highly similar at the nucleotide level. The deletion occurred within a 7-bp motif present at each deletion point in full-length elements (fig. 4). Translations revealed that the reading frame was maintained in two of these copies. Using inverse PCR, we determined that all three copies were inserted at the same genomic position. Thus, these copies result from vertical transmission of an ancestral deleted copy.

Inactive copies of the second type were identified in subfamilies E and K (RL4-110, M11-6, Ci-16, MK-28, and L15-16). These elements appeared to have been extensively mutated relative to other copies of the same subfamily. This high level of polymorphism, reflected by the long branches of the tree, appeared to be primarily due to CG-to-TA transitions (table 2), reminiscent of the repeat induced point mutation (RIP) process first discovered in *Neurospora crassa* (Selker 1987). An analysis of the 3′ and 5′ sequence context of C-T mutations (including G-A mutations, counted as C-T mutations on the complementary strand) revealed that mutations occurred primarily into \((\text{A/T/C})\text{pCp(A/G)}\), with A or T preferred in the 5′ flank and A preferred over G in the 3′ flank (table 2). We used a “de-RIP” approach (Cambareri et al. 1998) consisting of removal of all observed CG-to-TA transitions to investigate whether similarities of these copies to members of their families could be improved. In phylogenetic trees including these “de-RIPped” sequences, RL4-110(dRIP) grouped with the small clade containing *impE*, Ln88-23, S47-35, R-8, and Ln3-1 (99.7% identity with *impE*). M11-6(dRIP) and Ci-16(dRIP) remained more related to their “ripped” relatives (data not shown). Although we previously reported the existence of a “ripped” copy (*impB*) in strain M24 (Hua-Van et al. 1998), the pattern of transition was quite different between *impB*, RL4-110, and M11-6, all belonging to the E subfamily. In contrast, the patterns of mutation in the three “ripped” elements belonging to the K subfamily—Ci-16, MK-28, and L15-16—were very similar (no more than 2% nucleotide divergence among these elements). We determined that these three copies were inserted at the same genomic position and thus derived from one ancestral copy in which RIP occurred.

Genomic Positions of *impala* Copies

Using PCR, we investigated the presence of *impala* elements at the same genomic position in the strains previously used for PCR-detection of *impala*. Primers annealing to flanking regions were used for the detection of *impA*, *impC*, *impD*, and *impE*. For the 3′ deleted copy of *impB*, a primer hybridizing in the 5′ flanking region and an *impala*-specific internal primer (SpE3) were used. For the detection of the *impG* element, we chose a 3′ *impala* internal primer and a 5′ primer nested in solo-LTR *Han* that is inserted in the 5′ end of *impG* (Hua-Van et al. 1998, 2001). *Impala* elements were detected in eight of the 20 strains analyzed, and these strains have various host specificities (see table 3). *ImpA* was never detected via PCR; a short fragment corresponding to the empty site was always amplified. Absence of *impE* was also associated with the amplification of an empty site. In contrast, empty sites were never
amplified with the impD primers, and the impD element was detected only in strain M7. Complete absence of amplicons may be interpreted as the absence of one or both sequences complementary to the primers or an organization of the region different from that present in M24, or it could be due to the absence of the entire region. ImpB was amplified in four strains. The 3’ end of this copy was truncated in M24. In order to determine if impB was also truncated in these strains, we performed a PCR using a primer lying in the deleted part of impB. No amplification was obtained, suggesting that impB-homologous elements also possess 3’ deletions (data not shown). Moreover, the sequencing of the different amplified fragments of impB revealed the same mutational pattern observed in M24 impB (these sequences, B-impB, V-impB, and D-impB, were included in the tree shown in fig. 3).

Phylogenetic Relationships Among Strains

During the course of this study, we found several situations in which strains with different host specificities contained elements with similar mutations or elements that were present at the same genomic position. The presence of such elements might reflect a close genetic relationship between the strains. These results raised questions about the phylogenetic relationships among strains of *F. oxysporum* responsible for diseases in different hosts. To better understand these relationships, a phylogenetic analysis based on DNA sequences of a portion of the EF1α gene was conducted, as in O’Donnell et al. (1998). The aligned sequences were analyzed by maximum parsimony and bootstrapping (fig. 5). The analysis showed the existence of three main clades. We found that except for MK, all strains containing one or several copies of *impala* present at the same position as in strain M24 were grouped in one clade (73% bootstrap). MK and L15, both containing similar deleted and highly mutated copies, appeared very close to each other (97% bootstrap). The two other strains containing either a mutated copy (Ci) or a deleted copy (RL28) arose in the same clade (86%) (fig. 5).

### Table 2

Percentages of CG-to-TA Transitions and Nucleotide Context in Degenerate Copies

<table>
<thead>
<tr>
<th>Copies</th>
<th>Flank of Mutated C</th>
<th>CG to TA</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>impA</td>
<td>-</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>impB</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>impC</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>impD</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>impE</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>impG</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
</tbody>
</table>

a Percentages are based on the total number of nucleotide differences.

b Data include only changes that are not adjacent to one another.

c Relative to impE.

d Relative to Ci-36. Copies L15-16 and MK-28 are not reported here because of their high similarity to Ci-16.

### Table 3

Presence of *impala* Copies at the Same Position as in M24 in Other Strains

<table>
<thead>
<tr>
<th>Impala</th>
<th>SFA5</th>
<th>SFB5</th>
<th>SFC5</th>
<th>SFD5</th>
<th>SFE5</th>
<th>SFG5</th>
</tr>
</thead>
<tbody>
<tr>
<td>impA</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>impB</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>impC</td>
<td>-</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>impD</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>impE</td>
<td>-</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>impG</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**Note:** + indicates presence of the copy (1.5-kb PCR product); – indicates absence of the copy (PCR product of 200 bp); 0 indicates absence of PCR product amplification; ? indicates that results are unclear due to aspecific amplifications.
Diversity Among Subfamilies

Potentially active elements (with no interruption in the sequenced coding frame) have been identified in subfamilies E, D, and F. Inactive elements were found in all subfamilies and represent the majority of the sequences. Inactivity mainly results from stop codons and frameshift mutations. However, we detected other alterations resulting from large deletions or a high rate of transitions. All of these data suggest that the *impala* family is genetically diverse.

Three elements, containing the same internal deletion, were identified in three strains with different host specificities. These elements were located at the same chromosomal position and clearly arose by vertical transmission from a common ancestor that contained this deleted copy. Location of the deletion between two repeats of 7 bp suggests mechanisms similar to those proposed for several other deleted transposable elements (O’Hare and Rubin 1983; Streck, MacGaffey, and Beckendorf 1986). Another is homologous recombination between short repeats, from one element or from copies on sister chromatids (Streck, MacGaffey, and Beckendorf 1986). A third mechanism is the gap repair mechanism that takes place after excision of the element (Kurkulos et al. 1994; Hsia and Schnable 1996; Rubin and Levy 1997).

Diversity among subfamilies is also illustrated by the identification of several highly mutated elements in the E and K subfamilies. Alterations (CG-to-TA transitions) resemble those associated with RIP, a mutational process described in *N. crassa* (Selker et al. 1987). This process acts on duplicated sequences in the same nucleus during the sexual phase. RIP-like processes have also been identified in other ascomycetous fungi, *Aspergillus fumigatus* (Neuviel et al. 1996), *M. grisea* (Nakayashiki et al. 1999), and *Podospora anserina* (Hamann, Feller, and Osiewacz 2000). *Aspergillus fumigatus*, like *F. oxysporum*, is an asexual fungus. Although the type of alteration is the same, the RIP processes in all of these species differ in target preference. In *N. crassa*, C-to-T transitions occur with decreasing frequencies in Cpa > Cpt > Gpc > Cpc (Cambareri et al. 1989). C-to-T transitions are found mainly in (A/T)pCp(A/T) in *M. grisea*. In *F. oxysporum* mutations into Cpa and Gpc were preferred, and we also observed in the 5’ flanks of mutated C’s a preference for A, T, and, to a lesser extent, C. The de-RIPped sequences confirmed that RL4-110 and M11-6 belong to the E subfamily (99.8% and 98% identity with *impE*, respectively). The de-RIPped copy Ci-16 (dRIP) still showed 7% nucleotide divergence from Ci-36, suggesting that Ci-16 is not a ripped version of Ci-36 and comes from another copy, not identified during this work. This putative element could be a divergent member of the K subfamily or a member of another new subfamily. Taking into account that subfamilies usually diverge from each other by 20%–30%, the three ripped elements can be considered for copies Cu-12 and Cu-15 (*forma specialis cubense*) and Ci-32 and Ci-172 (*forma specialis ciceri*).

**Discussion**

*Impala* Is an Ancient Component of the *F. oxysporum* Complex

Distribution of the *impala* family within the *F. oxysporum* complex as investigated by Southern blot and PCR revealed that *impala* elements were present in most of the strains we tested. *Impala* thus appears to be an ancient component of the *F. oxysporum* genome. Amplification using specific primers and DNA sequencing revealed that subfamilies that had diverged by 20%–30% were frequently present within the same genome. This supports the hypothesis of an ancient presence of the different subfamilies in the *F. oxysporum* genome. The coexistence of these subfamilies in several strains is consistent with the vertical transmission of these copies present in a common ancestor. These subfamilies are always represented by a small number of copies, suggesting that *impala* is not a very active family or that selection acts to eliminate additional copies. However, occasional activity is evidenced by the presence of several copies of the same subfamily in some strains, characterized by strain-specific polymorphism, as observed
to belong to the K subfamily. Whereas the distributions of RIP in copies impB, RL4-110, and M11-6 are different, impB and its homologs, as well as the three copies in the K subfamilies, share nearly identical mutational patterns. In these two cases, the mutational process appears to have occurred in a common ancestor of the strains, since all copies homologous to impB are located at the same genomic position, as are the three ripped copies of subfamily K. Ripped copies appear to be the only remnant of impala in strains M11 and RL4 (see hybridization results in fig. 2B). Strains M24 and Cl contain ripped and nonripped elements from the same subfamily. At least in these strains, the RIP-like process may have acted transiently, with a possible reinsertion of impala.

Impala Evolution and Phylogenetic Relationship Within the F. oxysporum Complex

The presence of highly similar impala elements in strains belonging to different formae specialae raised the question of the evolutionary origin of these strains. We constructed a phylogeny with the nuclear gene EF1a, chosen because of its demonstrated utility for phylogeny reconstruction within this complex (O’Donnell et al. 1998). We showed that except for strain MK, discussed below, strains containing impala elements at the same genomic position are genetically related on the basis of the EF1a gene tree. This result provides evidence that strains may be more closely related than assumed by their host specificities. For instance, the group of strains related to M24 exhibits pathogenic specificities to a wide range of plants. This suggests that pathogenicity shifts are relatively frequent and raises questions concerning the genetic bases of such a versatile pathogenic behavior. The polyphyletic origins of some formae specialae, notably of melonis (O’Donnell et al. 1998), were also confirmed in this study. Strain MK was previously found to be different from other melonis strains based on VCG and mtDNA haplotype analysis (Jacobson and Gordon 1990). We showed here that this strain, collected in Mexico, is actually closely related to strain L15, a pathogen of tomato collected in North Africa.

The phylogenetic data help clarify the evolutionary history of the impala family. The presence of copies at the same genomic location in closely related strains strongly suggests that these copies were inserted in a common ancestor. Absence of these copies in other related strains might then indicate that they were lost, either by excision (for impA and impE) or by deletion or rearrangement of the region (for impD). Stochastic loss is also suggested by the fact that some strains lack some subfamilies and by the fact that some rare strains appear to be devoid of impala. However, the existence of impala elements belonging to more widely divergent subfamilies is not completely excluded. Too-divergent copies would be undetectable in standard Southern experiments or by PCR amplification.

Another mechanism that may play a role in transposable element evolution is horizontal transfer. This is suggested for several transposable elements, with the most convincing examples being illustrated by the P element from Drosophila (Daniels et al. 1990; Clark and Kidwell 1997; Silva and Kidwell 2000) and by the widely spread mariner element (Lohe et al. 1995; Robertson and Lampe 1995b). Members of this family have been found to be more similar among distantly related species than within a species. In the absence of strong phylogenetic support, horizontal transfer is not easily demonstrated. The widespread distribution of the different subfamilies is likely the result of an ancestral polymorphism. The horizontal transfer of one or more subfamilies is not excluded, but investigations on closely related species will be required for assessment of this question. Within the F. oxysporum complex, the distribution of different elements (ripped, deleted, or at a given position) is also suggestive of a vertical transmission, except in the case of impG in strain MK. This strain does not appear to be closely related to the other melonis strains on the basis of the EF1a analysis. Nevertheless, it contains the copy impG, detected only in a small group of melonis strains. Moreover, this strain contains a molecular marker (the sfo gene, homologous to the yeast SUR1 gene; Hua-Van et al. 2000), which has thus far been detected only in melonis strains (unpublished data). This gene is absent in all non-melonis strains, even those closely related to M24, and it is located 10 kb upstream of impG in M24. The presence of this gene and impG in MK suggests that the whole region could be conserved between MK and M24. However, this region, which is extremely rich in repeated sequences, is characterized by several apparent recombination events and is apparently subject to rapid reorganization (Hua-Van et al. 2000). Therefore, the possibility that the entire region surrounding impG in M24 could be present in MK seems rather surprising and raises the possibility of the genetic transfer of a large piece of DNA from the M24 group to MK. However, further analyses are required to test this hypothesis.

Impala as a Marker of Genetic Relationships

Identification of closely related elements (by their sequences or by their genomic positions) in different strains raises the question of whether impala could be used as a marker for inferring genetic relationships. The study of specific fixed impala elements appears to be a good strategy to detect close relationships. However, the example of strain MK indicates that some exceptions can occur. Therefore, the use of impala as a marker for genetic relationships can be very helpful in detecting unsuspected relationships, but it has to be used along with, or confirmed by, other markers, such as single-copy nuclear genes.

Acknowledgments

We are grateful to K. O’Donnell for helpful comments on the manuscript. We thank J. M. Daviere for his help in sequencing, and C. Gerlinger for her technical assistance.
LITERATURE CITED


Hsie, A. P., and P. S. Schinabe. 1996. DNA sequence analyses support the role of interrupted gap repair in the origin of internal deletions of the maize transposon, MuDR. Genetics 142:603–618.


Evolution of impala Transposable Elements 1969


Thomas Eickbush, reviewing editor

Accepted June 20, 2001