Genome Sequencing of the Plant Pathogen Taphrina deformans, the Causal Agent of Peach Leaf Curl

Cissé, Ousmane H.

2013-03-30


http://hdl.handle.net/10138/166983
https://doi.org/10.1128/mBio.00055-13

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
RESEARCH ARTICLE

Genome Sequencing of the Plant Pathogen *Taphrina deformans*, the Causal Agent of Peach Leaf Curls

Ousmane H. Cissé, João M. G. C. F. Almeida, Álvaro Fonseca, Ajay Anand Kumar, Jarkko Salojärvi, Kirk Overmyer, Philippe M. Hauser, Marco Pagni

Institute of Microbiology, University Hospital and University of Lausanne, Lausanne, Switzerland; Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; Centro de Recursos Microbiológicos (CREMI), Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; Division of Plant Biology, Department of Biosciences, University of Helsinki, Helsinki, Finland

P.M.H. and M.P. contributed equally to the work.

ABSTRACT *Taphrina deformans* is a fungus responsible for peach leaf curl, an important plant disease. It is phylogenetically assigned to the Taphrinomycotina subphylum, which includes the fission yeast and the mammalian pathogens of the genus *Pneumocystis*. We describe here the genome of *T. deformans* in the light of its dual plant-saprophytic/plant-parasitic lifestyle. The 13.3-Mb genome contains few identifiable repeated elements (ca. 1.5%) and a relatively high GC content (49.5%). A total of 5,735 protein-coding genes were identified, among which 83% share similarities with other fungi. Adaptation to the plant host seems reflected in the genome, since the genome carries genes involved in plant cell wall degradation (e.g., cellulases and cutinases), secondary metabolism, the hallmark glyoxylate cycle, detoxification, and sterol biosynthesis, as well as genes involved in the biosynthesis of plant hormones. Genes involved in lipid metabolism may play a role in its virulence. Several locus candidates for putative MAT cassettes and sex-related genes akin to those of *Schizosaccharomyces pombe* were identified. A mating-type-switching mechanism similar to that found in ascomycetous yeasts could be in effect. Taken together, the findings are consistent with the alternate saprophytic and parasitic-pathogenic lifestyles of *T. deformans*.

IMPORTANCE Peach leaf curl is an important plant disease which causes significant losses of fruit production. We report here the genome sequence of the causative agent of the disease, the fungus *Taphrina deformans*. The genome carries characteristic genes that are important for the plant infection process. These include (i) proteases that allow degradation of the plant tissues; (ii) secondary metabolites which are products favoring interaction of the fungus with the environment, including the host; (iii) hormones that are responsible for the symptom of severely distorted leaves on the host; and (iv) drug detoxification enzymes that confer resistance to fungicides. The availability of the genome allows the design of new drug targets as well as the elaboration of specific management strategies to fight the disease.

Taphrina deformans is the agent of peach leaf curl, a disease that affects orchards throughout the temperate regions of the world. Hosts include peach (*Prunus persica*) and, to a lesser extent, almond trees (*Prunus dulcis*). *T. deformans* is the most intensively studied species of the genus due to its importance as a plant pathogen worldwide, all known peach and nectarine cultivars being susceptible to infection (1). The economic burden of peach leaf curl is variable, depending on the areas of production. The disease causes $2.5 to $3 million in losses in the United States annually (2). In northern Italy, it represents an important threat to the tree and causes $2.5 to $3 million losses per year (3). One of the most significant symptomatic effects occurs on leaves that become severely distorted, reddish (4). These symptoms are usually attributed to the production of indole-3-acetic acid (IAA) by *T. deformans* (5).

*T. deformans* is a fungus belonging to the subphylum Taphrinomycotina, a basal lineage of the phylum Ascomycota. The genus *Taphrina* belongs to the order Taphriniales in the class Taphrinomycetes. The Taphrinomycotina subphylum also contains the fission yeasts of the genus *Schizosaccharomyces* and the animal parasites of the genus *Pneumocystis* (6). *T. deformans* is dimorphic, i.e., it has a biotrophic hyphal state which is parasitic-pathogenic on plant tissues and a saprophytic yeast state that may overwinter on plant surfaces. Only the yeast state can be grown on conventional culture media in vitro (7).

We report here the sequence of the genome of *T. deformans* strain PYCC 5710, as well as comparative genomic analysis of its gene repertoire. The genome carries key genes for plant invasion and colonization.

RESULTS Nuclear and mitochondrial genomes. The *T. deformans* strain PYCC 5710 genomic DNA sequencing yielded a 13.3-Mb genome assembly with a 29.3-fold coverage (Table 1). This corresponds to 394 scaffolds, 50% of which are longer than 72 kb. The complete-
ness of the genome was estimated to be 98.7%, and 99% of the 3,923 publicly available expressed sequence tags (ESTs) could be mapped on the assembly. A single copy of the nuclear rRNA operon is present in a single scaffold (scaffold 361). The 18S rRNA gene displays 100% identity with the previously reported allele of this gene (8), confirming the identity of the strain sequenced here. An additional truncated nonfunctional cox1 gene contains two group I introns (15), the cox2 gene contains one group II intron (16), and the cox2 gene contains a group II intron (17). The T. deformans mitochondrial genome carries an additional truncated nonfunctional cox1 gene. The same feature has been reported in Ustilago maydis (J. C. Kennell and C. Bohmer, unpublished data), but the biological implication of this is not known.

**Gene content.** We identified 5,735 protein-coding genes, 44% of which are supported by ESTs and 83% by similarity to proteins found in other fungi (see Table S4 in the supplemental material). About 30% of the total genes were manually curated, including all those which are discussed in detail in the text. To better understand the gene content of T. deformans, we compared its proteome to those of 16 relevant fungi, including two plant-nonpathogenic Taphrinomycotina members, seven plant pathogens, four plant saprophytes, two human pathogens, and one animal pathogen. These fungi were selected to reflect different host infection mechanisms. The maximum likelihood phylogeny derived from alignment of 227 single-copy orthologs is consistent with previous phylogenies (6, 18, 19), placing T. deformans in the basal Taphrinomycotina subphyllum (Fig. 1A).

Ortholog clustering identified 13,566 protein groups containing at least one protein from T. deformans (Fig. 1B). Out of the 5,735 T. deformans proteins, 22.3% are included in clusters present in all species (shown as green strips in Fig. 1B); 58.4% are shared, i.e., present in T. deformans and at least one other species (shown as blue strips); and 4.4% are present only in Taphrinomycotina members (shown as orange strips).

To identify genes that might be responsible for T. deformans pathogenicity, we classified organisms according to their lifestyle. Fungi can be free-living as saprophytes recycling nutrients and/or can establish interactions with a host (20). These interactions are diverse and can vary from biotrophy (i.e., the fungus is able to live within its host and cannot be cultured in vitro in the case of obligate biotrophy) to necrotrophy (i.e., the fungus kills its host to access nutrients). Here, we deliberately simplified the problem by considering an organism to be a pathogen if it can cause plant disease and to be a saprophyte if it can live on dead plant material.

### Table 1: T. deformans nuclear genome statistics and comparison to those of other members of the Taphrinomycotina subphyllum

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>T. deformans</th>
<th>P. jirovecii&lt;sup&gt;a&lt;/sup&gt;</th>
<th>S. pombe&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly size (Mb)</td>
<td>13.3</td>
<td>8.1</td>
<td>12.5</td>
</tr>
<tr>
<td>Average GC (%)</td>
<td>49.5</td>
<td>29.1</td>
<td>36.0</td>
</tr>
<tr>
<td>No. of scaffolds</td>
<td>394</td>
<td>356&lt;sup&gt;c&lt;/sup&gt;</td>
<td>39&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Scaffold N&lt;sub&gt;D&lt;/sub&gt; (kb)</td>
<td>71.9</td>
<td>41.6&lt;sup&gt;d&lt;/sup&gt;</td>
<td>NA&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Telomere repeat motif</td>
<td>TTAGGG</td>
<td>TTAGGG</td>
<td>TTAC(A)(C)(G)(1–8)</td>
</tr>
<tr>
<td>Repeat (%)</td>
<td>1.5&lt;sup&gt;f&lt;/sup&gt;</td>
<td>9.8</td>
<td>9.4</td>
</tr>
<tr>
<td>No. of protein-coding genes</td>
<td>5,735</td>
<td>3,898</td>
<td>5,124</td>
</tr>
<tr>
<td>Gene density (genes per Mb)</td>
<td>431</td>
<td>481</td>
<td>554</td>
</tr>
<tr>
<td>Average gene length (nt)</td>
<td>1,492</td>
<td>1,472</td>
<td>1,446</td>
</tr>
<tr>
<td>Mean exon per gene</td>
<td>2.1</td>
<td>3.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Mean intron length (nt)</td>
<td>78</td>
<td>61</td>
<td>81</td>
</tr>
<tr>
<td>No. of tRNAs</td>
<td>169</td>
<td>71</td>
<td>171</td>
</tr>
</tbody>
</table>

<sup>a</sup> These data are from reference 18. P. jirovecii infects specifically humans.

<sup>b</sup> These statistics are from http://www.pombase.org (73).

<sup>c</sup> Contigs.

<sup>d</sup> Chromosomes.

<sup>e</sup> NA, not applicable or not available from the website

<sup>f</sup> Detailed description of T. deformans repeated elements is shown in Table S2 in the supplemental material.

<sup>g</sup> nt, nucleotide.
Host-pathogen interaction. During its life cycle, *T. deformans* enters and damages plant tissues. The genome encodes a full repertoire of carbohydrate-active enzymes, including those required for degrading the plant cell wall (see Table S7 in the supplemental material). Additionally, the genome encodes 70 proteases, among which 21 appear to be secreted (see Table S8 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phaus-mpdbdata.htm). Although *T. deformans* possesses cellulases and xylanases that allow the degradation of cellulose and hemicellulose, it lacks lignin-degrading enzymes necessary for the degradation of lignocellulose and lignin components (FOLymes). These enzymes are also absent in other fungi (i.e., *U. maydis*, Cryt-tococcus neoformans, *N. crassa*, and Saccharomyces cerevisiae) (25). Other plant fungal parasites, such as the white rot fungus *Schizophyllum commune*, possess a full repertoire of FOLymes (25). After entering into its host organism, *T. deformans* is able to cause stable infection using different mechanisms and effectors. Among them, the most important are the following.

(i) Plant hormones and carotenoids. The production of the plant auxin hormone indole-3-acetic acid (IAA) by *T. deformans* is generally considered to be responsible for the formation of characteristic leaf deformation symptoms (5, 26). In *U. maydis*, the tryptophan aminotransferases (encoded by *Tam* genes) and indole-3-acetaldehyde dehydrogenases (encoded by *Iad* genes) are required for production of IAA (27). As in *U. maydis*, two orthologs of each *Tam* gene (locus tags TAPDE_000671 and TAPDE_001394) are present in *T. deformans* and at least one other species (blue), Taphrinomycotina specific (orange), present in *T. deformans* and plant saprophytes (yellow), present in *T. deformans* and plant fungal pathogens (red), other proteins (dark gray), and species specific (light gray).

The production of the auxin hormone induces characteristic leaf deformation symptoms in *T. deformans*. This is achieved through the production of IAA biosynthesis genes (5). Two orthologs of each *Tam* gene are present, along with at least one other species, Taphrinomycotina specific, present in *T. deformans* and plant saprophytes, plant fungal pathogens, and species specific. The enzymes encoded by these genes are crucial for plant-fungus interaction (23). The production of IAA is essential for pathogenesis, as it allows the fungus to enter and damage plant tissues. The genome encodes a full repertoire of FOLymes, including cellulases and xylanases, which are necessary for the degradation of lignocellulose and lignin components. Additionally, the genome encodes 70 proteases, among which 21 appear to be secreted. These enzymes are absent in other fungi, such as *U. maydis*, Cryp- tococcus neoformans, *N. crassa*, and Saccharomyces cerevisiae. Other plant fungal parasites, such as the white rot fungus *Schizophyllum commune*, possess a full repertoire of FOLymes. After entering into its host organism, *T. deformans* is able to cause stable infection using different mechanisms and effectors. Among them, the most important are the following.

(i) Plant hormones and carotenoids. The production of the plant auxin hormone indole-3-acetic acid (IAA) by *T. deformans* is generally considered to be responsible for the formation of characteristic leaf deformation symptoms (5, 26). In *U. maydis*, the tryptophan aminotransferases (encoded by *Tam* genes) and indole-3-acetaldehyde dehydrogenases (encoded by *Iad* genes) are required for production of IAA (27). As in *U. maydis*, two orthologs of each *Tam* gene (locus tags TAPDE_000671 and TAPDE_001394) are present in *T. deformans* and at least one other species (blue), Taphrinomycotina specific (orange), present in *T. deformans* and plant saprophytes (yellow), present in *T. deformans* and plant fungal pathogens (red), other proteins (dark gray), and species specific (light gray).

The production of the auxin hormone induces characteristic leaf deformation symptoms in *T. deformans*. This is achieved through the production of IAA biosynthesis genes (5). Two orthologs of each *Tam* gene are present, along with at least one other species, Taphrinomycotina specific, present in *T. deformans* and plant saprophytes, plant fungal pathogens, and species specific. The enzymes encoded by these genes are crucial for plant-fungus interaction (23). The production of IAA is essential for pathogenesis, as it allows the fungus to enter and damage plant tissues. The genome encodes a full repertoire of FOLymes, including cellulases and xylanases, which are necessary for the degradation of lignocellulose and lignin components. Additionally, the genome encodes 70 proteases, among which 21 appear to be secreted. These enzymes are absent in other fungi, such as *U. maydis*, Cryp- tococcus neoformans, *N. crassa*, and Saccharomyces cerevisiae. Other plant fungal parasites, such as the white rot fungus *Schizophyllum commune*, possess a full repertoire of FOLymes. After entering into its host organism, *T. deformans* is able to cause stable infection using different mechanisms and effectors. Among them, the most important are the following.

(i) Plant hormones and carotenoids. The production of the plant auxin hormone indole-3-acetic acid (IAA) by *T. deformans* is generally considered to be responsible for the formation of characteristic leaf deformation symptoms (5, 26). In *U. maydis*, the tryptophan aminotransferases (encoded by *Tam* genes) and indole-3-acetaldehyde dehydrogenases (encoded by *Iad* genes) are required for production of IAA (27). As in *U. maydis*, two orthologs of each *Tam* gene (locus tags TAPDE_000671 and TAPDE_001394) are present in *T. deformans* and at least one other species (blue), Taphrinomycotina specific (orange), present in *T. deformans* and plant saprophytes (yellow), present in *T. deformans* and plant fungal pathogens (red), other proteins (dark gray), and species specific (light gray).

The production of the auxin hormone induces characteristic leaf deformation symptoms in *T. deformans*. This is achieved through the production of IAA biosynthesis genes (5). Two orthologs of each *Tam* gene are present, along with at least one other species, Taphrinomycotina specific, present in *T. deformans* and plant saprophytes, plant fungal pathogens, and species specific. The enzymes encoded by these genes are crucial for plant-fungus interaction (23). The production of IAA is essential for pathogenesis, as it allows the fungus to enter and damage plant tissues. The genome encodes a full repertoire of FOLymes, including cellulases and xylanases, which are necessary for the degradation of lignocellulose and lignin components. Additionally, the genome encodes 70 proteases, among which 21 appear to be secreted. These enzymes are absent in other fungi, such as *U. maydis*, Cryp- tococcus neoformans, *N. crassa*, and Saccharomyces cerevisiae. Other plant fungal parasites, such as the white rot fungus *Schizophyllum commune*, possess a full repertoire of FOLymes. After entering into its host organism, *T. deformans* is able to cause stable infection using different mechanisms and effectors. Among them, the most important are the following.

(i) Plant hormones and carotenoids. The production of the plant auxin hormone indole-3-acetic acid (IAA) by *T. deformans* is generally considered to be responsible for the formation of characteristic leaf deformation symptoms (5, 26). In *U. maydis*, the tryptophan aminotransferases (encoded by *Tam* genes) and indole-3-acetaldehyde dehydrogenases (encoded by *Iad* genes) are required for production of IAA (27). As in *U. maydis*, two orthologs of each *Tam* gene (locus tags TAPDE_000671 and TAPDE_001394) are present in *T. deformans* and at least one other species (blue), Taphrinomycotina specific (orange), present in *T. deformans* and plant saprophytes (yellow), present in *T. deformans* and plant fungal pathogens (red), other proteins (dark gray), and species specific (light gray).

The production of the auxin hormone induces characteristic leaf deformation symptoms in *T. deformans*. This is achieved through the production of IAA biosynthesis genes (5). Two orthologs of each *Tam* gene are present, along with at least one other species, Taphrinomycotina specific, present in *T. deformans* and plant saprophytes, plant fungal pathogens, and species specific. The enzymes encoded by these genes are crucial for plant-fungus interaction (23). The production of IAA is essential for pathogenesis, as it allows the fungus to enter and damage plant tissues. The genome encodes a full repertoire of FOLymes, including cellulases and xylanases, which are necessary for the degradation of lignocellulose and lignin components. Additionally, the genome encodes 70 proteases, among which 21 appear to be secreted. These enzymes are absent in other fungi, such as *U. maydis*, Cryp- tococcus neoformans, *N. crassa*, and Saccharomyces cerevisiae. Other plant fungal parasites, such as the white rot fungus *Schizophyllum commune*, possess a full repertoire of FOLymes. After entering into its host organism, *T. deformans* is able to cause stable infection using different mechanisms and effectors. Among them, the most important are the following.

(i) Plant hormones and carotenoids. The production of the plant auxin hormone indole-3-acetic acid (IAA) by *T. deformans* is generally considered to be responsible for the formation of characteristic leaf deformation symptoms (5, 26). In *U. maydis*, the tryptophan aminotransferases (encoded by *Tam* genes) and indole-3-acetaldehyde dehydrogenases (encoded by *Iad* genes) are required for production of IAA (27). As in *U. maydis*, two orthologs of each *Tam* gene (locus tags TAPDE_000671 and TAPDE_001394) are present in *T. deformans* and at least one other species (blue), Taphrinomycotina specific (orange), present in *T. deformans* and plant saprophytes (yellow), present in *T. deformans* and plant fungal pathogens (red), other proteins (dark gray), and species specific (light gray).
mans strain PYCC 5710 sequenced here is among the highest carotenoid accumulators among Taphrinomycotina members (28). Accordingly, all genes necessary for carotenoid biosynthesis are present in the assembly, including the geranylgeranylpyrophosphate synthase (locus tag TAPDE_002204), the lycopene cyclase/phytoene synthase (locus tag TAPDE_005571), and the phytoene dehydrogenase (locus tag TAPDE_002896). Carotenoids are important protective compounds in fungi, safeguarding against UV light and other oxidative stresses (29). This role may be of particular importance in T. deformans, which lives exposed in the high-light environment at or near the plant surface. In U. maydis, carotenoids can also function as precursors for the production of retinal, a molecule which acts as a chromophore in the phototactic protein opsin (29). Similarly, the genome of T. deformans carries three opsin genes (locus tags TAPDE_003195, TAPDE_001673, and TAPDE_005609), as well as a single copy of the carotenoid cleavage oxygenase (locus tag TAPDE_002691), an enzyme which cleaves β-carotene into two molecules of opsin. The role of carotenoids and their derivatives in stress biology and photobiology is not characterized in T. deformans. In plants, carotenoids are important as precursors of the plant stress hormone abscisic acid (ABA). In T. deformans, we identified genes orthologous to those involved in ABA biosynthesis of the plant parasite fungus Botrytis cinerea, which are distinct from those of plants (30). The ABA biosynthesis cluster includes a cytochrome P450 monooxygenase similar to B. cinerea BAB1 and BAB2 (locus tag TAPDE_000660), an ortholog of BAB3 (locus tag TAPDE_004842), and several short-chain dehydrogenases/reductases similar to BAB4 (locus tags TAPDE_000889, TAPDE_001904, TAPDE_002246, TAPDE_002541, and TAPDE_000083). Additionally, the T. deformans genome harbors putative orthologs of the plant ABA biosynthesis genes, i.e., zeaxanthin epoxidase (locus tag TAPDE_003610), xanthoxin dehydrogenase (locus tag TAPDE_000917), molybdenum cofactor sulforase (locus tag TAPDE_004842), 9-cis-epoxy-carotenoid dioxygenase (locus tag TAPDE_002691), and abciscic-aldehyde oxidase (locus tag TAPDE_004279). These findings suggest the possibility of ABA production, which has not been previously reported or tested for in T. deformans, although observed in the ascomycetous plant pathogen B. cinerea (30). However, expression studies would be required to determine if these genes are activated upon stress or under other conditions. Our phylogenetic analyses did not provide conclusive evidence of the transfer of the plant-like ABA biosynthesis gene cluster from plant to T. deformans (data not shown).

(ii) Mating genes. In the T. deformans genome, several locus candidates for putative MAT cassettes akin to those of S. pombe were identified, namely, potential orthologs of matnc and matpi genes, as well as of genes involved in mating type switching (e.g., swl1, swl5, swl6, and swl10). Many other genes described as sex related in S. pombe seem to be also present (31), e.g., those encoding the components of the pheromone-activated mitogen-activated protein kinase (MAPK) signaling cascade (see Table S9 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauser-suppldata.htm). The genome of T. deformans was found to be mating-type switching (ap1, swl2, and msh3), meiosis (e.g., mei2), or silencing (clr1) and one of the pheromone receptor genes (mam2) are either absent or appear to be truncated. Others, like a putative dcr1 homolog, lack relevant domains, in this case the helicase at the N terminus, but conserve all the others.

(iii) Secondary metabolism. Secondary metabolites are small active molecules that are produced by many microorganisms, including bacteria, fungi, and plants. They are usually not essential but can confer advantages when microbes are interacting with the environment (32). Four genes that define putative secondary metabolite biosynthesis gene clusters have been identified in the T. deformans genome. Three of them bear a similarity to nonribosomal peptide synthases (NRPS), and one appears to be a hybrid NRPS/polyketide synthase (PKS) (see Table S10 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauser-suppldata.htm). Since these biosynthetic gene clusters are not present in other Taphrinomycotina members (i.e., Schizosaccharomyces spp. and Pneumocystis spp.), T. deformans appears to have a unique capacity for the production of secondary metabolites.

Drug targets. The genome of T. deformans encodes targets for known antifungal drugs. Despite its apparent resistance toazole treatment (33), we identified the gene encoding the lanosterol 14-alpha-demethylase which is targeted by the phytosanitary antifungal molecules difenoconazole, tebuconazole, and myclobutanil (34, 35). Moreover, most of the genes involved in the biosynthesis of different sterols (brassicasterol, ergosterol, and cholesterol) (see Table S11 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauser-suppldata.htm) are conserved. The apparent resistance to azoles might be explained by the fact that T. deformans uses brassicasterol as a major sterol in its membranes, instead of ergosterol (36). Genes encoding β,1,3 glucan synthase, which is targeted by the antifungal echinocandins, and cytochrome bc1, which is targeted by tri-foxystrobin, were also identified. Additionally, T. deformans possesses a homolog of the Bcm1p1 gene, which was shown to be crucial for the resistance of Botryotinia fuckeliana to the fungicides fludioxonil, camptothecin, and cercosporin (37).

Metabolic capabilities. The T. deformans yeast state is able to grow on conventional fungal or yeast media (7). Consistently, its genome encodes most of the respiratory and biosynthetic pathways (e.g., citric acid cycle, glycolysis, and amino acid synthesis). Nevertheless, several genes involved in the metabolism of galactose and other sugars are lacking (see Table S12 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauser-suppldata.htm). This is consistent with the fact that T. deformans has a relatively narrow spectrum of assimilation of carbon and nitrogen sources. Indeed, sugars utilized include cellobiose, D-xylene, and D-glucose but not galactose, sucrose, maltose, or sorbose, or trehalose (7). Additionally, T. deformans has been reported to be auxotrophic for thiamine and biotin (38), and consistently, genes specific to the biosynthesis of these vitamins are missing in the genome (see Table S12 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauser-suppldata.htm). These biosynthetic gene clusters have been identified in the T. deformans genome. Three of them bear a similarity to nonribosomal peptide synthases (NRPS), and one appears to be a hybrid NRPS/polyketide synthase (PKS) (see Table S10 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauser-suppldata.htm). Since these biosynthetic gene clusters are not present in other Taphrinomycotina members (i.e., Schizosaccharomyces spp. and Pneumocystis spp.), T. deformans appears to have a unique capacity for the production of secondary metabolites.
DISCUSSION

The \textit{T. deformans} nuclear genome is larger and harbors a larger number of genes than those of \textit{P. jirovecii} and \textit{S. pombe}. Its mitochondrial genome is similar in size to that of \textit{P. jirovecii} but larger than that of \textit{S. pombe}. The nuclear genome presents a low content of repeated elements and the same telomere repeat as that in plants, mammals, and filamentous fungi (18, 40). The apparently low content of transposable elements seems not to be correlated with RIP activity, which is also absent in other Taphrinomycotina members. A single copy of the ribosomal DNA unit has been identified in the genome. To date, this characteristic was observed only in \textit{Pneumocystis} species (41, 42); it is generally believed to be associated with a low replication rate and is compatible with the notably slow growth of the yeast state of \textit{T. deformans} in culture (7).

The genome also carries genes for important mechanisms such as respiratory and biosynthetic pathways, plant cell and tissue degradation, sterol biosynthesis, and mating genes. This gene content allows this organism to be both a plant pathogen and saprophyte.

Despite its relatively reduced genome compared to those of other fungal plant pathogens (e.g., 13 versus 59.9 Mb for \textit{Fusarium oxysporum}), about 5% of \textit{T. deformans} genes are conserved only in fungal plant pathogens. These genes are likely to constitute the genetic basis of its virulence. Enrichment analysis indicated that many of these genes are involved in cutinase activity, lipid metabolism, and response to oxidative stress activity, which are all key factors for successful invasion of plant cells (21, 22). The cutinases hydrolyze the cutin, which protects the surface of the plant leaves. Consequently, cutinases are predicted to be involved in the early phase of the infection. The lipid metabolism is crucial for the maintenance of cell wall integrity and the signaling network and thus might be important for virulence. The genes involved in the response to oxidative stress might also provide protection against stress resulting from exposure to high-intensity sunlight and/or plant defense mechanisms.

Two genomic features make \textit{T. deformans} unique among Taphrinomycotina members. First, it harbors four putative secondary metabolite gene clusters, although such clusters were lost during the evolution of other ascomycetes such as \textit{S. pombe} (43), \textit{S. cerevisiae} (43), and \textit{P. jirovecii} (18). However, their functionality and final products in \textit{T. deformans} remain to be determined. Second, the genome carries genes putatively involved in the biosynthesis of plant hormones (IAA) that may be responsible for the formation of characteristic leaf deformation symptoms. It has been suggested that hormone production capabilities as well as secondary metabolite gene clusters in other fungi could have been acquired via horizontal gene transfer (44, 45). However, the hormone biosynthetic pathways differed between fungi and plants in all cases examined so far (30, 46). This is consistent with the view that plant hormone biosynthesis results rather from a convergent evolution involving different metabolic pathways or enzymes, which produce similar compounds in plants and fungi (47).

This fungus is homothallic since single haploid cells grow rise to dikaryotic hyphae and asci on plant tissue (48). In contrast to \textit{U. maydis}, conjugation of cells heterozygous at the mating type (MAT) locus does not occur. However, since heterozygosity of paired nuclei in the dikaryotic hyphae is often required for pathogenic development, a mating-type-switching mechanism similar to that found in ascomycetous yeasts could be in effect. Many locus candidates for the expected sex-related genes and MAT cas-settes were identified in the \textit{T. deformans} genome. The MAT cas-settes are notoriously difficult to characterize due to the short lengths of their coding sequences (CDSs) and their low degree of sequence conservation. Two of the three putative cas-settes seem incomplete, lacking one of the expected genes, and some genes involved in mating type switching are either absent or truncated. The missing cassette and genes may have gone undetected or are present in gaps of the present genome sequence.

MATERIALS AND METHODS

Strain and culture conditions. The \textit{T. deformans} strain PYCC 5710 (=CBS 356.35) cells were grown in liquid YEPD medium (1% [wt/vol] Difco yeast extract, 2% Difco peptone, 2% glucose) at 25°C with rotary shaking (150 × g). Cells were harvested by centrifugation at 6,000 rpm for 1 min, washed twice with 30 ml of EDTA (20 mM, pH 5.8), aliquoted (5 g per aliquot), and stored at −20°C until use.

Genomic DNA isolation, sequencing, and assembly. Genomic DNA from a single aliquot was isolated using the protocol of Lachance (49) and used for whole-shotgun sequencing. Data were generated by a combination of both mate-paired and single-end Roche 454 GS FLX sequencing with Titanium chemistry. Raw data statistics can be found in Table S13 at http://www.chuv.ch/imu/imu_home/imu_recherche/imu_recherche_hauser/imu-phauer-suppldata.htm. Vectors and other contaminants (7.3% of total reads) were removed using Phred and Cross match processing tools (50, 51) with the UniVec database (available at ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec/) and a custom collection of bacterial genomes. The mitochondrial genome reads were isolated (ca. 4.3% of total reads), and assembled separately (see Text S2 in the supplemental material). The nuclear genome was assembled using the remaining filtered sequences (88.4%) with Newbler (version 2.6). The genome completeness was estimated using the CEGMA pipeline (52) and alignment of 3,923 ESTs (BLASTn cutoff, ≤10−5).

Gene prediction and functional annotations. The \textit{ab initio} gene finder GeneMark-ES (53) was self-trained on the genome sequence assembly. Augustus (54) and SNAP (55) gene finders were trained on 3,923 expressed sequence tags (ESTs) available at http://www.ncbi.nlm.nih.gov/nucest/?term=\textit{Taphrina\%20deformans}. Final gene models were produced using Maker pipeline (v.1.21) (56), and Rfam (v.10.1) (61). Mapping of coding sequences (CDSs) to KEGG biochemical pathways was performed using Priam (62).

Gene family analysis and phylogeny. Gene families for \textit{T. deformans} were built using OrthoMCL with a Markov inflation of 1.5 and a maximum E value of 10−5 (69). The fungal proteomes used were those of \textit{S. pombe}, \textit{Magnaporthe oryzae}, \textit{Fusarium oxysporum}, \textit{Botrytis cinerea}, \textit{Puccinia graminis}, \textit{Ustilago maydis}, \textit{Rhizopus oryzae}, \textit{Neuro-
spora crassa, Aspergillus oryzae, Sordaria macrospora, Uncinocarpus resii, Pyrenophora tritic-repentis, Trichophyton rubrum, Coccidioides posadasi, Batrachochytrium dendrobatidis (available at http://www.uniprot.org); release 2012_09, P. jirovecii (18), and T. deformans (this study). Single-copy orthologs were identified using OMA (standalone version 0.99q) (70), concatenated, and aligned using MAFFT with the L-INS-i method (71). Phylology was inferred using RAxML (v.7.2.8) (72) with 100 bootstrap replicates and BLOSUM62 as model.

**Additional supplemental material.** Additional supplemental material can be found at http://www.chuv.ch/imul/imu_home/imu_recherche /imu_recherche_hauser/imu-phauser-suppldata.htm.

**Nucleotide sequence accession numbers.** The whole-genome sequence has been deposited at ENA-EMBL under accession no. CAHR00000000. Raw reads have been deposited at ENA SRA under accession no. ERP001279.

**SUPPLEMENTAL MATERIAL**


**Text S1, DOCX file, 0.1 MB.**

**Text S2, DOCX file, 0.1 MB.**

**Figure S1, TIF file, 0.5 MB.**

**Table S1, DOCX file, 0.1 MB.**

**Table S2, DOCX file, 0.1 MB.**

**Table S6, DOCX file, 0.1 MB.**

**ACKNOWLEDGMENTS**

This work was supported by the Swiss National Science Foundation grant 310030-124998 to P.M.H. and M.P. and by an Academy of Finland Fellowship to K.O. (decision no. 251397 and 256073). The computations were performed at the Vital-IT Center for high-performance computing (http://www.vital-it.ch) of the SIB (Swiss Institute of Bioinformatics). SIB receives financial supports from the Swiss Federal Government through the State Secretariat for Education and Research (SER).

We thank Thomas Bernard for help with enzyme annotation and helpful discussions. We thank Cláudia Carvalho (PYCC, Portugal) for preparing genomic DNA.

**REFERENCES**


9. Keely SP, Renaud H, Wakefield AE, Cushion MT, Smulian AG, Fonser


15. Lang BF. 1984. The mitochondrial genome of the fission yeast Schizosaccharomyces pombe: highly homologous introns are inserted at the same position of the otherwise less conserved cox1 genes in Schizosaccharomyces pombe and Aspergillus nidulans. EMBO J. 3:2129–2136.


---

Cissé et al.