The mitochondrial genome of the phytopathogenic basidiomycete Moniliophthora perniciosa is 109 kb in size and contains a stable integrated plasmid


Laboratório de Genômica e Expressão – Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, 13083-970, Campinas – SP, Brazil
Laboratório de Ecotoxicologia, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, 13400-970, Piracicaba – SP, Brazil
Laboratório de Bioinformática da Universidade Estadual de Santa Cruz, 45650-000, Ilhéus – BA, Brazil
Laboratório de Pesquisa em Microbiologia (LAPEM), Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana (UEFS), 44031-460, Feira de Santana – BA, Brazil
Instituto de Computacão, Universidade Estadual de Campinas, 13084-971, Campinas – SP, Brazil
Ludwig Institute For Cancer Research, 01509-010, São Paulo – SP, Brazil
Departamento de Genética e Evolução e Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética – CBMEG, Universidade Estadual de Campinas, 13035-875, Campinas – SP, Brazil
Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas, 13083-970, Campinas – SP, Brazil
CEPLAC/CEPEC/SFIT, 45600-970, Itabuna – BA, Brazil
Laboratório de Biologia Molecular – Faculdade de Farmácia, Universidade Federal da Bahia, 40170-290, Salvador – BA, Brazil
Laboratório de Biologia Molecular – Departamento de Biologia Geral, Instituto de Biologia, Universidade Federal da Bahia, 40170-290, Salvador – BA, Brazil
Sustainable Perennial Crops Laboratory, USDA-ARS, 10300 Baltimore, Av., BARC-W, Beltsville, MD 20740, USA
Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, 45650-000, Ilhéus – BA, Brazil

We present here the sequence of the mitochondrial genome of the basidiomycete phytopathogenic hemibiotrophic fungus Moniliophthora perniciosa, causal agent of the Witches’ Broom Disease in Theobroma cacao. The DNA is a circular molecule of 109103 base pairs, with 31.9 % GC, and is the largest sequenced so far. This size is due essentially to the presence of numerous non-conserved hypothetical ORFs. It contains the 14 genes coding for proteins involved in the oxidative phosphorylation, the two rRNA genes, one ORF coding for a ribosomal protein (rps3), and a set of 26 tRNA genes that recognize codons for all
Introduction

Moniliophthora perniciosa is the causal agent of witches’ broom disease (WBD) in cacao (Theobroma cacao). This fungus is a basidiomycete previously classified as Crinipellis perniciosa, and recently, it was reclassified due to its close relation to Moniliophthora roreri (Aime & Phillips-Mora 2005; Evans et al. 2002), the causal agent of frosty pod rot (FPR) in cacao. These two represent the two most devastating cacao diseases in the Americas. M. perniciosa presents a hemibiotrophic life style, with differentiated mycelia in each phase. The biotrophic mycelia are mononucleated, grow very slowly and colonize the intercellular spaces (apoplast) of the infected plant tissue. This phase is difficult to obtain in the laboratory, but recently we have characterized the conditions for its ex planta growth (Meinhardt et al. 2006). The saprophytic mycelia are dikaryotic with clamp connections and grow vigorously in axenic cultures.

Due to the economical importance, our group is involved in the study of the complex biology of the hemibiotrophic fungus M. perniciosa (Garcia et al. 2007; Meinhardt et al. 2006; Rincones et al. 2003; Rincones et al. 2006; Rincones et al. in press; Rio et al. 2008; Scarpari et al. 2005). Included in this investigation is a genome project (http://www.lge.ibi.unicamp.br/vassoura), and during the course of the sequencing we obtained the complete mitochondrial genome.

Besides M. perniciosa, 49 fungal mitochondrial genomes have been completely sequenced and annotated to date (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/futax_short.html). These genomes encompass species from all phyla. Despite the importance of the Basidiomycota, only five mitochondrial genomes from this group have been studied.

Mitochondria are generally accepted as vestigial endosymbiotic proteobacterial ancestors (Gray et al. 1999; Lang et al. 1999), and they can evolve faster than nuclear DNA (Burger et al. 2003; Burger & Lang 2003; Nosek & Tomaska 2003). The relationship between mitochondria and the nucleus has evolved over time and several mitochondrial genes encoding proteins for mitochondrial metabolism have been transferred to the nuclear genome (Adams & Palmer 2003; Burger et al. 2003; Dimmer & Rapaport 2008). Fungal mitochondrial genomes are smaller than those of plants, but larger than animal mitochondrial genomes (Burger et al. 2003). The gene content of the typical fungal mitochondrial genome is greatly reduced (ca 30–40 genes) in comparison to the genomes of bacteria and the mitochondrial genomes of other eukaryotes. Some conserved characteristics of mitochondrial genomes are the high A + T content, lack of methylation, conservation of the gene function, and high copy number (Campbell et al. 1999).

Fungal mitochondrial genomes vary greatly in size, show different topology, and even in the use of the genetic code among species (Gray et al. 1999). The number of genes can fluctuate enormously from three to 67 for protein encoding genes and from seven to 27 for tRNAs (Adams & Palmer 2003). They usually contain 14 genes encoding hydrophobic subunits of respiratory chain complexes, two genes for the large and small ribosomal subunits and a set of tRNAs (Gray et al. 1999). Differences in length and organization of the intergenic regions, differences in intron content (from 0 to 30) and size (ranging from 0.15 to 4 kb), unidentified ORFs, and the presence of plasmids contribute significantly to the size variability of the genomes among species (Bertrand 2000; Burger et al. 2003; Clark-Walker 1992; Hur et al. 1997). Also, there is no apparent correlation between size and gene content in the mitochondrial genomes (Burger et al. 2003).

The shortage of reliable morphological and ultrastructural characters is a challenge in the taxonomic classification of fungi. However, in recent years the use of mitochondrial sequence data of proteins involved in essential processes conserved throughout evolution have been used routinely to build phylogenetic trees for the characterization of the origin and evolution of fungal organisms (Bullerwell et al. 2003; Bullerwell & Lang 2005; Seif et al. 2005). Therefore, more complete mtDNA sequences from Basidiomycota may provide insights to the origin and evolution of fungal mitochondria.

Plasmids have been found in many different fungi (Griffiths 1995; Kempken 1995a; Meinhardt et al. 1990, 1997). All of the plasmids found in these fungi are mitochondrial with the possible exception of the Alternaria plasmids (Shepherd 1992). Plasmids are of two types, circular and linear. In Neurospora spp., both circular and linear types are found. The plasmids of Absidia glauca are circular (Haenfler et al. 1992) and all the remaining examples are linear. Some features are common to almost all of the mitochondrial linear plasmids, namely: (1) terminal inverted repeats; (2) proteins covalently attached to the 5’ termini; and (3) open reading frames (ORFs) for viral-like RNA and DNA polymerases (Meinhardt et al. 1990, 1997). Thus, these linear plasmids in filamentous fungi are thought to be descended from a common ancestor. However, horizontal transfer of a linear plasmid in Ascomycota is also shown (Kempken 1995b).

Although in most cases no function has been assigned for these elements, in some species of filamentous fungi their
presence and activity interferes with the growth of mycelia (Bertrand et al. 1985, 1986). In Neurospora spp. insertion of the plasmid DNA into the mitochondrial genome results to the disruption of several genes leading to senescence and death due to defective respiration (Bertrand 2000; Griffiths 1992, 1998; Rieck et al. 1982). Conversely, in Podospora anserina strain AL2 and in the slime mould Physarum polycephalum, which are species that normally senesce, insertion of the plasmids into the mtDNA has been correlated with an increased life-span (Hermanns et al. 1994; Maas et al. 2007; Nakagawa et al. 1998). Among Basidiomycota, invertron-like DNA plasmids have been completely sequenced in Agaricus bitorquis (Robison & Horgen 1999), Flammulina velutipes (Nakai et al. 2000), and Pleurotus ostreatus (Kim et al. 2000), but no specific function has been associated with their presence in these hosts.

The present work is a comprehensive study of the M. perniciosa mitochondrial genome. We describe its organization, gene content, and order, together with a comparative and phylogenetic analysis with other sequenced mitochondrial genomes. We provide evidence for the presence of potentially new mitochondrial genes and for the evolution of the mitochondrial genome by the stable integration of a linear mitochondrial plasmid.

### Material and methods

#### Fungal isolate, library construction and sequence assembly

Total DNA was extracted from strain CP02 of Moniliophthora perniciosa, as described previously (Rincones et al. 2003). We observed a relation between the integrity of the isolated DNA and the age of the culture used for the extraction (the older the plate, the worst the quality of the DNA). For that reason, a total of 50 genomic libraries were constructed using DNA isolated from young and old mycelia. Fragments ranging in size from 1–2 and 2–4 kbp were obtained from total DNA by sonication or nebulization (Surzicki 2000), and then cloned into the SmaI site of pUC18 or pCR4Blunt (TOPO Shotgun Subcloning kit, Invitrogen - Life Technologies, São Paulo, Brasil). Each library corresponded to independent cloning events using DNA obtained from individually growing cultures of CP02. The inserts were sequenced and analysed following the whole genome shotgun (WGS) approach (Venter et al. 1998).

Reads were assembled using the software package Phred/Phrap/Consed (Ewing & Green 1998; Ewing et al. 1998; Gordon et al. 1998) and the accuracy of the assembly was confirmed with the program CAP3 (Huang & Madan 1999).

All isolates used in this study can be obtained from the Laboratório de Genômica e Expressão at UNICAMP, as well as further information concerning the collection data.

#### Gene annotation

For the sake of clarity, we have defined the ORFs found in the mitochondrial genome as follows: (1) conserved known genes, as those normally present in mitochondrial genomes with known function; (2) conserved hypothetical ORFs, as those ORFs present also in other mitochondrial genomes, but without a defined function; and (3) non-conserved hypothetical ORFs, as those ORFs that do not overlap with conserved known genes or conserved hypothetical ORFs, and coding for products with minimum size compatible with the smallest conserved known genes (atp8 has 52 aa, atp9 has 73 aa and nad4L has 87 aa).

Conserved known genes have been named as their counterparts in other genomes. Intrinsic conserved hypothetical ORFs have been named as ci###gene. Intrinsic non-conserved hypothetical ORFs have been named as ci###gene. Conserved hypothetical ORFs located in the plasmid region have been named hypP#. Non-conserved hypothetical ORFs have been named as hyp#. Here # represents a number and gene is the name of the gene where it its located.

Conserved known genes and conserved hypothetical ORFs were detected by similarity searches for known genes and ORFs present in other fungal mitochondrial genomes using BLAST (Altschul et al. 1990). Non-conserved hypothetical ORFs were detected using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/).

Regions coding for tRNAs were identified with the program tRNAscan-SE (Lowe & Eddy 1997). The rRNAs sequences were identified by comparison with homologous Basidiomycota counterparts. Repeats were analysed using Reputer (Kurtz et al. 2001) and Tandem Repeats Finder (Benson 1999).

Exon/intron boundaries of conserved known genes were identified by comparison and alignment, using BLAST and CLUSTAL (Higgins 1994), with the genes from other fungal mitochondrial genomes.

#### PCR amplification

Amplification of selected regions of the mitochondrial genome was carried out by PCR. Primers P1 (5’-GCA GGGAAGGGATATATAGG-3’), P2 (5’-TTTGGAGAGCATCAAATCC-3’), P3 (5’-TTTTGAGAGACCATCAAATCC-3’), and P4 (5’-AA AGAACCTGAAATCCGG-3’) were designed to mach regions outside (P1 and P4) and inside (P2 and P3) of the inverted repeats that we did use to define the boundaries of the inserted plasmid (see Results). Total DNA from isolates in Table 2 purified as described above was used for the PCR experiments.

#### GC content, GC skew, and codon usage

GC content and GC skew (Grigoriev 1998) were calculated using the program Analyse seq (http://ludwig-sun2.unil.ch/~vioannid/TP_module5/18-03/freak2.html). Local GC content was calculated with sliding windows over 5000 bp and 500 bp with a pace of 500 and 50 bp, respectively. Codon usage was analysed using the program CODONW (Peden 1999).

#### Principal component analysis (PCA)

A matrix presenting the relative use of each codon for each ORF was evaluated by PCA in order to search for patterns of codon usage (Joliffe 1986). These data were submitted to analysis by the method of singular value decomposition of non-centred and non-scaled data matrix by means of the R software (R Development Core Team 2007). The number of principal components to be retained was determined by
and *Flammulina velutipes* copy of the plasmid (analysis. Three of the species used present more than one netic tree. A total of 13 complete plasmids were used in this study (both genes in the opposite orientation flanked by inverted repeats) were constructed with the program PhyML (Guindon & Gascuel 2003) using the ML method, with JTT substitution matrix and gamma distribution. The value of the shape parameter $\alpha$ used was optimized by the program PhyML. Branch support was tested by BS analysis with 100 resamplings (Felsenstein 1985). 

Genes that code for the mitochondrial cytochrome c oxidase (*cox1*, *cox2* and *cox3*), properly annotated, from 50 fungi with complete sequence for the mitochondrial genome were used to build the phylogenetic tree. The optimized value of the shape parameter $\alpha$ was 1.444 for this set of sequences.

Also, the genes *dpoB* and *rpo* from the mitochondrial plasmid from ten species that present a complete invertron structure (both genes in the opposite orientation flanked by inverted repeats) (Griffiths 1995) were used to build a phylogenetic tree. A total of 13 complete plasmids were used in this analysis. Three of the species used present more than one copy of the plasmid (*Neurospora crassa*, *Pleurotus ostreatus*, and *Flammulina velutipes*). Sequences with similarity to *dpoB* or *rpo* not organized with an invertron structure were not included in this analysis. The optimized value of the shape parameter $\alpha$ was 1.549 for this set of sequences.

**Comparative genomics**

The comparison of the gene content and order of all known mitochondrial genomes of Basidiomycota was carried out by visual inspection due to the small number of genomes used, six. All conserved known genes coding for proteins and the rRNAs were designed as boxes linearly as they appear in the genome beginning with the large rRNA (ml) in a clockwise direction.

**Results**

**Assembly of the mitochondrial genome**

After the initial assembly of 124,565 reads produced in the sequencing of all the libraries a large contig enclosing 5448 reads was generated. This contig was identified as the mitochondrial genome of the fungus by comparison to other fungal mitochondrial genomes. Inclusion of additional reads from new sequencing rounds and a careful assembly resulted in a final contig containing 6920 reads. The consensus sequence corresponded to a circular DNA of 109,103 bp without gaps, regions of low consensus quality, or regions of high-quality discrepancy (Phrap assembly). This corresponded to an approximate coverage of 31 times. The complete mtDNA sequence is available at GenBank (accession number AY376688).

Although most fungal mitochondrial genomes are circular (Torriani et al. 2008; Wang et al. 2008; Zivanovic et al. 2005), some linear genomes have been reported (Forget et al. 2002; Rycovska et al. 2004). Due to the programs we use not being able to assemble circular DNAs we have used a different strategy in order to check whether the mitochondrial genome of *M. perniciosa* is linear or circular. The strategy consisted in the masking of unique regions (minimum length of 750 bp) of the genome and new assemblies of all the reads were performed with the exception of the masked region. This was repeated masking different regions independently. If the mitochondrial genome is linear this strategy should render at least two different contigs representing the regions on both sides of the masked region. The result was always one contig comprising all the genome except the masked region; thus indicating that the mitochondrial genome is circular.

**Conserved known genes**

Fig 1 shows the map of the 119 possible coding regions present in the mitochondrial genome. Seventeen genes were classified as conserved known genes, coding for proteins were identified by comparison with other fungal mitochondrial genomes where the function of those have been described. Forty-one of those are involved in the oxidative and energy metabolism; seven subunits of the NADH dependent dehydrogenases, three cytochrome c oxidases, three F0 subunits of the ATP synthases and one cytochrome b. One gene, *rps3*, is involved in ribosome assembly (Bullerwell et al. 2000). The last two code for polymerases, a DNA-dependent RNA polymerase (*rpo*) and a DNA-directed DNA polymerase (*dpoB*).

Two rRNAs and 26 tRNAs were identified in the mitochondrial genome. These tRNAs are grouped into regions of the genome presenting a relatively high GC content (Fig 3). The location of the tRNAs is shown at the most inner circle in Fig 1, and Table 1 list all the tRNAs anti-codons with its corresponding amino acid.

**Conserved and non-conserved hypothetical ORFs**

Twelve additional putative ORFs have been defined as conserved hypothetical ORFs based on the fact of the existence of similar sequences in other fungal mitochondrial genomes. Three of them are present between a set of inverted repeats together with the genes *dpoB* and *rpo*, and the other nine are intronic ORFs located inside the genes that code for *cox1*, *cob*, *nad4* and *nad5* (Fig 1).

The analysis of the remainder of the sequence with the program ORF Finder located 59 possible non-conserved hypothetical ORFs and three non-conserved intronic ORFs located inside the genes *nad5*, *cob* and *cox2* (Fig 1). One copy of the LAGLIDADG motif was found in one hypothetical ORF, hyp15.

**Introns**

Twelve introns have been identified: *cox1* has six introns, *cox2* and *cob* have two introns each, and both *nad4* and *nad5* have one intron each (Fig 1). All the six introns in *cox1*, one of the introns in *cob*, and the intron in *nad4* and *nad5* contain one conserved hypothetical ORF each. Besides these conserved
hypothetical ORFs, cob, cox2, and nad5 contain one non-conserved hypothetical ORF each.

Two different motifs belonging to two types of homing endonucleases have been found in eight of the intronic conserved hypothetical ORFs. Two copies of the LAGLIDADG motif were found in six of them (oi1cox1, oi2cox1, oi4cox1, oi5cox1, oi1nad5, and oi1nad4), and oi1cob presented only one of such motifs (Heath et al. 1997). Inside the intronic ORF oi6cox1 one GIY-YIG motif was found (van Roey et al. 2001).

Codon usage and ORF orientation

The preferential codon usage was calculated from the exon sequences of the conserved known genes (Table 1). At least one tRNA for every amino acid has been identified, and for one of them it was not possible to determine the anti-codon. Most of the codons can be translated with tRNAs present in the mitochondrial genome by the wobble rule (Jukes 1984). There are two codons, AUA and UGA, without the corresponding tRNA.

When the PCA was applied to all the potential protein-coding genes (91 known and hypothetical ORFs) a clear pattern could be identified (Fig 2). The genes segregated into two different groups (Fig 2, left panel). One first group (A) was formed up by 45 non-conserved hypothetical ORFs, and the three intronic non-conserved hypothetical ORFs: oBi1nad5, oBi1cob, and oBi1cox2; and a second group (B) made up by the remaining 43 ORFs; including all conserved known genes (17), nine conserved hypothetical ORFs, and 14 non-conserved hypothetical ORFs. It is likely that these non-conserved, hypothetical ORFs represent true genes due to the fact that they have the same preferential codon usage and the same clockwise orientation as the conserved known genes. Only five ORFs are found in the
anticlockwise direction; two conserved known genes (atp8 and dpoB), one conserved hypothetical ORF (hypP3), and two non-conserved hypothetical ORFs. These three hypothetical ORFs are located immediately upstream of the gene dpoB. Also, all the tRNAs and the two rRNAs are in the clockwise orientation.

**GC content**

The mitochondrial genome presents a relatively low GC content (31.9%) in comparison to that of the genomic DNA (47.7%), calculated using all the non-mitochondrial contigs. Comparing the GC average for positions one (GC1), two (GC2), and three (GC3) of conserved protein coding genes and for all mtDNA (GCm), we found that GC3 was significantly biased towards lower values (supplementary data), which has been observed in protein coding genes in some mitochondrial genomes (Stewart & Beckenbach 2006).

The GC content and the cumulative GC skew of the mitochondrial genome are shown in Fig 3. The GC content (Fig 3, blue and black lines) presents some variation along the mitochondrial genome with three well-defined significant regions. It begins with a content between 35 and 40 %, maintaining this

### Table 1 – Codon usage in the conserved known genes from the mitochondrial genome

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>E</th>
<th>Anti-codon</th>
<th>Amino acid</th>
<th>Codon</th>
<th>E</th>
<th>Anti-codon</th>
<th>Amino acid</th>
<th>Codon</th>
<th>E</th>
<th>Anti-codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>334</td>
<td>(g)aa</td>
<td>Ser</td>
<td>UCU</td>
<td>167</td>
<td>(u)ga</td>
<td>Tyr</td>
<td>UAU</td>
<td>217</td>
<td>(g)ua</td>
</tr>
<tr>
<td></td>
<td>UUC</td>
<td>96</td>
<td>GAA</td>
<td></td>
<td>UCC</td>
<td>16</td>
<td>(u)ga</td>
<td></td>
<td>UAC</td>
<td>39</td>
<td>GU</td>
</tr>
<tr>
<td>Leu</td>
<td>UUA</td>
<td>545</td>
<td>UAA</td>
<td></td>
<td>UCA</td>
<td>140</td>
<td>UGA</td>
<td>Ter</td>
<td>UAA</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UUG</td>
<td>52</td>
<td>(u)aa</td>
<td></td>
<td>UCG</td>
<td>9</td>
<td>(u)ga</td>
<td></td>
<td>UAG</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CUU</td>
<td>79</td>
<td>(u)ag</td>
<td>Pro</td>
<td>CCU</td>
<td>109</td>
<td>(u)gg</td>
<td>His</td>
<td>CAU</td>
<td>64</td>
<td>(g)ug</td>
</tr>
<tr>
<td></td>
<td>CUC</td>
<td>5</td>
<td>(u)ag</td>
<td></td>
<td>CCC</td>
<td>3</td>
<td>(u)gg</td>
<td></td>
<td>CAC</td>
<td>36</td>
<td>GUG</td>
</tr>
<tr>
<td></td>
<td>CUA</td>
<td>72</td>
<td>UAG</td>
<td></td>
<td>CCA</td>
<td>47</td>
<td>UGG</td>
<td>Gln</td>
<td>CAA</td>
<td>111</td>
<td>UUG</td>
</tr>
<tr>
<td></td>
<td>CUG</td>
<td>21</td>
<td>(u)ag</td>
<td></td>
<td>CCG</td>
<td>3</td>
<td>(u)gg</td>
<td></td>
<td>CAG</td>
<td>14</td>
<td>(u)ug</td>
</tr>
<tr>
<td>Ile</td>
<td>AUA</td>
<td>270</td>
<td>(g)au</td>
<td>Thr</td>
<td>ACU</td>
<td>135</td>
<td>(u)gu</td>
<td>Asn</td>
<td>AUA</td>
<td>235</td>
<td>(g)au</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>44</td>
<td>GAU</td>
<td></td>
<td>ACC</td>
<td>5</td>
<td>(u)gu</td>
<td></td>
<td>AAC</td>
<td>34</td>
<td>GUU</td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>102</td>
<td>CAU</td>
<td></td>
<td>ACG</td>
<td>4</td>
<td>(u)gu</td>
<td></td>
<td>AAG</td>
<td>18</td>
<td>(u)uu</td>
</tr>
<tr>
<td></td>
<td>GUU</td>
<td>135</td>
<td>(u)ac</td>
<td>Ala</td>
<td>GCU</td>
<td>164</td>
<td>(u)gc</td>
<td>Asp</td>
<td>GAU</td>
<td>119</td>
<td>(g)ac</td>
</tr>
<tr>
<td></td>
<td>GUC</td>
<td>5</td>
<td>(u)ac</td>
<td></td>
<td>GCC</td>
<td>13</td>
<td>(u)gc</td>
<td></td>
<td>GAC</td>
<td>20</td>
<td>GUC</td>
</tr>
<tr>
<td></td>
<td>GUA</td>
<td>160</td>
<td>UAC</td>
<td></td>
<td>GCA</td>
<td>91</td>
<td>UGC</td>
<td>Glu</td>
<td>GAA</td>
<td>111</td>
<td>UUC</td>
</tr>
<tr>
<td></td>
<td>GUG</td>
<td>20</td>
<td>(u)ac</td>
<td></td>
<td>GCG</td>
<td>13</td>
<td>(u)gc</td>
<td></td>
<td>GAG</td>
<td>16</td>
<td>(u)uc</td>
</tr>
</tbody>
</table>

The tRNAs that can potentially match codons by the wobble rule are indicated by small caps in parentheses. Missing tRNAs are indicated by a dash, and the stop codons are indicated by asterisks.

**Fig 2 – PCA of codon usage.** Left panel shows a plot of the sequence distribution according to codon usage (two-component solution accounting for 78.1 % of the total variance; 65.4 and 12.7 % from components one and two, respectively). Right panel shows the relations of ORFs classified by group and orientation.

**Group A**

CWO - oB1cox, oB1cox2, hyp5, hyp6, hyp12, hyp17, hyp19, hyp21, hyp22, hyp23, hyp24, hyp26, hyp27, hyp31, hyp32, hyp37, hyp38, hyp40, hyp41, hyp42, hyp43, hyp44, hyp47, hyp48, hyp49, hyp51, hyp53, hyp54, hyp55, hyp56

CCWO - oB1nad5, hyp1, hyp2, hyp3, hyp4, hyp16, hyp18, hyp25, hyp30, hyp33, hyp34, hyp35, hyp39, hyp45, hyp46, hyp50, hyp52, hyp59

**Group B**

CWO - atp6, atp9, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, rpo, rps3, oB1cox, oB1cox1, oB1nad4, oB1nad5, oB2cox1, oB3cox1, oB4cox1, oB5cox1, oB6cox1, hypP1, hypP2, hyp7, hyp8, hyp9, hyp10, hyp11, hyp13, hyp14, hyp15, hyp20, hyp28, hyp29, hyp36

CCWO - atp8, dpoB, hypP3, hyp57, hyp58
average up till the 19 kbp position where it begins to drop down to around 24 % at the 28 kbp position; then it begins to raise up to around the previous level at the 45 kbp position just under the 30 % content; then it goes up to its maximum local content just over 40 % between the 55 to the 59 kbp position; then the content goes back to the 35 % average content up to the 71 kbp position where it drops down, for the last time, to its lower value (around 23 %) at the 76 kbp position; it remains at this low value up till the 88 kbp position where it raises again up to an average content of 35 % at the 92 kbp position, and remains around this value till the end.

The cumulative GC skew (Fig 3, red line) raises up to a maximum at the 42 kbp position and then drops down to the 72 kbp position where it goes up again until generating a new local maximum between the 55 to the 59 kbp position; after this position levels up and its relatively constant up till the end of the sequence.

The most significant feature of this analysis comes from the combined observation of the GC content and the GC skew. The region located between the 73 kbp and the 90 kbp positions (Fig 3, between dotted lines 1 and 4) is characterized by the lowest GC content and a change in the direction of the cumulative GC skew (red line). Also conspicuous is the significant reduction in the noise of the profile of the GC content calculated with the smaller window (500 bp) of this region with respect to the rest of the genome (blue line). This kind of pattern has been interpreted as regions of genome rearrangement (Grigoriev 1998). This region contains nine ORFs and four sets of inverted repeats. It begins with one set of inverted repeats 347 bp long (Fig 3, dotted line 1), continues with two conserved hypothetical ORFs (hypP1 and hypP2), then a set of inverted repeats 130 bp long followed by rpo and dpoB, another set of inverted repeats 130 bp long followed by one conserved hypothetical ORF (hypP3), then the last set of inverted repeats (Fig 3, dotted line 2) appears followed by two non-conserved hypothetical ORFs (hypP8 and hypP7) with a anticlockwise orientation (Fig 3, dotted line 3), and then the region ends with two more non-conserved, hypothetical ORFs (hyp28 and hyp29) with a clockwise orientation (Fig 3, dotted line 4). The region delimited by the two bigger sets of inverted repeats that contains the genes rpo and dpoB, together with three conserved hypothetical ORFs (Fig 3, dotted lines 1 and 2), resembles the structure of a linear plasmid (Griffiths 1995).

The analysis of the GC skew in mammals (Touchon et al. 2005) and bacteria (Arakawa et al. 2007) has proved useful in the determination of putative origin and termini of replication, indicated by a change in the polarity of this index. Despite not having any experimental data about the location of the origin of replication, we are proposing that the putative origin of replication of the mitochondrial genome of Moniliovithora perniciosa might be located around position 1 where we have a minimum for the GC skew index (Fig 3).

Integrated plasmid

The analysis of the sequence showed the presence of two genes, rpo and dpoB, that code for a DNA-dependent RNA
polymerase and a DNA-directed DNA polymerase, respectively. These two genes are not typical mitochondrial genes. They are found in plasmids, and these plasmids can be free in the linear form or integrated into the mitochondrial genome (Griffiths 1995, 1998).

Four sets of inverted repeats were found in the region surrounding these two genes, two small, 130 bp long, and two large, 347 bp long. The organization of this region is described above and shown in Fig 4. Invertron-type plasmids are characterized by large terminal inverted repeats of around 1000 bp (Griffiths 1998), as also found in transposable elements, bacteriophages, and adenoviruses (Sakaguchi 1990). We are proposing the existence of an invertron-type plasmid integrated in the Moniliophthora perniciosa mitochondrial genome in the region delimited by the large inverted repeats (347 bp).

The plasmid region presents three conserved, hypothetical ORFs; two of them (hypP1 and hypP2) are located before rpo and have the same orientation as this polymerase. The other ORF, hypP3, is located after dpoB and it has the same orientation as this gene, but opposite to the first three genes (Fig 4A). These two sets of conserved, hypothetical ORFs are located between the large and the small inverted repeats. Four additional, non-conserved, hypothetical ORFs (hyp58, hyp57, hyp28, and hyp29) are located after the last large inverted repeat. These ORFs are inside the region of low GC content and with a significant deviation of the GC skew (Fig 3, between dotted lines 2 and 4). The first two (hyp58 and hyp57) have the same orientation as dpoB, and the last two (hyp28 and hyp29) have the opposite orientation. A comparison of the protein sequence of these hypothetical ORFs showed that hypP1, hypP3, and hyp58 are significantly similar (BlastX e-values: hypP1 and hyp58 – 1e-19 and hypP1 and hypP3 – 2e-8); as it happens with hyp57 and hyp28 (BlastX e-values: hyp57 and hyp28 – 7e-15). The orientation and similarity among these sequences seems to indicate the integration and duplication of the genes present at this location.

In order to check the relationship between the plasmid and the mitochondrial genome, we followed two different strategies. The first one consisted in a careful analysis of the assembly of the sequencing data, and the second one involved an analysis of the sequence by PCR. The analysis of the assembly of the sequencing data did not show any overrepresentation of reads along the whole mtDNA assembly, and more specifically at the plasmid region. In the shotgun strategy used here an overrepresentation of reads is an indication of the presence of a different number of copies of some part of the DNA. This should indicate the presence of independent entities, like plasmids. This is the case of the mitochondrial genome that is overrepresented with respect to the genomic DNA, 31 versus 2.5 times, respectively. Another indication of the existence of a separate plasmid would be the abrupt interruption of the reads at the ends of free plasmids, but this situation was not detected. This in silico analysis is a strong indication that the plasmid did not exist in an independent form. Therefore, it may be stably integrated in the genome.

It is important to note that genomic libraries were constructed from mycelia at different ages, with young and old mycelia. It has been reported that some linear plasmids play a role in fungal aging, normally called senescence, by jumping over the mtDNA presenting different insertion points in the sequence, disrupting this way essential genes (D’Souza et al. 2005; Fox & Kennell 2001). If the aging of M. perniciosa mycelia is caused by a similar process, we should be able to detected different combinations of mitochondrial and plasmid DNAs.

**Fig 4** – (A) Genetic map of the plasmid and adjacent regions. Black boxes represent the two polymerases. Conserved hypothetical ORFs, hypP1 and hypP3, together with non-conserved hypothetical ORFs hyp58 that show significant similarity (BlastX e-values: hypP1 and hyp58 – 1e-19 and hypP1 and hypP3 – 2e-8) are represented by light grey boxes. Non-conserved hypothetical ORFs, hyp57 and hyp28, also show significant similarity (BlastX e-values: hyp57 and hyp28 – 7e-15) are represented by dark grey boxes. Inverted repeats are represented by the small black boxes connected by a line according to its size. The position of the primers P1, P2, P3, and P4 are indicated by the boxes on top, as well as the expected size of the amplification product. (B) Analysis of the boundaries of the plasmid region by PCR using primers P1–P2 (1–2) and P3–P4 (3–4). Origin of the samples: BA, Bahia; RO, Roraima; PA, Para. Dashed lines indicate the same regions as in Fig 3.
by misalignments in the assembly. However, such inconsistencies, which would be detected by ‘high-quality discrepancies’ (Ewing & Green 1998; Ewing et al. 1998), have not been detected and the assembly consistently produced the same consensus sequence. Therefore, we concluded that the plasmid was stably integrated in a defined position in the mtDNA of this isolate.

An experimental approach was also devised in order to further check the integration of the plasmid DNA in other isolates. Four different primers were designed, two of them just outside the plasmid sequence representing sequences belonging to the mtDNA, and the other two inside the boundaries of the region considered as being the plasmid representing sequences at regions outside the larger external inverted repeats. The region to be amplified by these primers encompasses the external repeats that define the boundaries of the plasmid. A representation of the position of the primers and the expected size of the amplified fragment is shown in Fig 4A. In the case of the plasmid being integrated, fragments corresponding to the integration boundaries will be amplified. The results of the PCR analysis performed with five isolates from different regions of Brazil are shown in Fig 4B. The PCR analysis was repeated with further 23 isolates of M. perniciosa. The nomination and origin of all isolates used in this study are summarized in Table 2. In all isolates we detected the amplification product in the conditions used here.

The results obtained with the genes encoding proteins of the respiratory enzymes (nad1–nad5, nad4–nad6) are shown in Figs 4 and 5. Comparison of the mitochondrial genomes of Moniliophthora perniciosa with that of Podospora anserina because they are two of the biggest, over 100 kb, described so far (Fig 5). Interestingly, the main reason for the large size of the P. anserina mitochondrial genome is the presence of numerous introns (Cummings et al. 1990), whereas the mitochondrial genome from M. perniciosa possesses many non-conserved hypothetical ORFs.

**Comparative genomics**

The inspection of the mitochondrial genomes of Moniliophthora perniciosa together with those from other Basidiomycota species did not show significant overall synteny of the protein coding genes (Fig 5). However, a more detailed inspection showed that nad genes are present as groups (nad2–nad3, nad4L–nad5, and nad4–nad6); there is a tendency to stay together on the genome in the three members of the class Agaricomycetes (Schizophyllum commune, Pleurotus ostreatus, and M. perniciosa).

We compared the mitochondrial genomes of M. perniciosa with that of Podospora anserina because they are two of the biggest, over 100 kb, described so far (Fig 5). Interestingly, the main reason for the large size of the P. anserina mitochondrial genome is the presence of numerous introns (Cummings et al. 1990), whereas the mitochondrial genome from M. perniciosa possesses many non-conserved hypothetical ORFs.

**Phylogeny**

Phylogenetic trees were built using the three cytochrome oxidase genes (cox1, cox2, and cox3) from 50 fungal species independently. The trees obtained for the individual genes were essentially the same, without significant differences. Due to this a new phylogenetic tree was built using the three cox genes concatenated. This tree showed the same topology as those built with independent genes, but presented higher BS support for most of the branches, with the exception of some basal branches. Fig 6 shows the rooted phylogenetic tree for this analysis of the three cox genes concatenated. The results largely agree with fungal taxonomy and with recent fungal phylogenies (James et al. 2006; Kouvelis et al. 2004).

The unrooted phylogenetic tree built with the polymerase genes from the plasmid using all complete fungal invertron-like plasmids, including kalilo-like plasmids, is shown in Fig 7. There is a clear separation of the plasmid sequences into two major groups. One of them was made up by the kalilo-like plasmids (Fig 7, right), as described by Sakaguchi (1990) and the other by the non-kalilo-like linear plasmids (Fig 7, left). This separation has no correlation with the phylogenetic separation between Basidiomycota and Ascomycota obtained with the genes encoding proteins of the respiratory

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Local of collection</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP02</td>
<td>Itabuna, BA, Brazil</td>
<td>UESC a</td>
</tr>
<tr>
<td>CP09</td>
<td>Ilhéus, BA, Brazil</td>
<td>CEPLAC b</td>
</tr>
<tr>
<td>Belmont</td>
<td>Belmonte, BA, Brazil</td>
<td>UFB c</td>
</tr>
<tr>
<td>Ilhéus</td>
<td>Ilhéus, BA, Brazil</td>
<td>UFB c</td>
</tr>
<tr>
<td>Santo Amaro</td>
<td>Santo Amaro, BA, Brazil</td>
<td>UFB</td>
</tr>
<tr>
<td>FA42, FA276, FA277, FA278, FA562, FA563</td>
<td>Itabuna BA, Brazil FAC d</td>
<td></td>
</tr>
<tr>
<td>FA281</td>
<td>Aiquara, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>FA287</td>
<td>Inema, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>FA293</td>
<td>Gandu, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>FA300</td>
<td>Itabira, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>FA311</td>
<td>Itagiba, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>FA317</td>
<td>Ilhéus, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>BP10</td>
<td>Itapebi, BA, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>FA551</td>
<td>Tabatinga, AM, Brazil</td>
<td>FAC e</td>
</tr>
<tr>
<td>ESJOH-1</td>
<td>Marituba, PA, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-2</td>
<td>Ouro Preto, RO, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-3</td>
<td>Belém, PA, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-4</td>
<td>Altamira, PA, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-5</td>
<td>Medicilandia, PA, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-6</td>
<td>Ariquemes, RO, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-7</td>
<td>Manaus, AM, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-8</td>
<td>Ji-Paraná, RO, Brazil</td>
<td>ESALQ f</td>
</tr>
<tr>
<td>ESJOH-9</td>
<td>Alta Floresta, MT, Brazil</td>
<td>ESALQ f</td>
</tr>
</tbody>
</table>

a UESC (Universidade Estadual de Santa Cruz) Ilhéus, Bahia, collected by Júlio Cascando.
b CEPLAC (Comissão Executiva do Plano de Lavoura do Cacau) Ilhéus, Bahia, collected by Karina Gramach.
c UFB (Universidade Federal de Brasília) Brasília, collected by Maricília Arruda.
d FAC (Fazenda Almirante Cacau) Ilhéus, Bahia, collected by Alan Pomett.
e ESALQ - CENA (Centro de Energia Nuclear na Agricultura, in Escola Superior de Agricultura “Luiz de Queiroz”) Piracicaba, São Paulo, collected by Paulo Albuquerque (ERJOH – CEPLAC).
chain (Fig 6). In this analysis, M. perniciosa genes grouped together with the kalilo-like plasmids from other fungi. There are other types of plasmids with a circular topology that preset only one of the polymerases (Griffiths 1995). These plasmids were not used in the phylogenetic analysis. Although they present sequence similarity, they represent non-homologous genes from different types of plasmids with different evolutionary origin (Griffiths 1995).

Morphology of Moniliophthora perniciosa in vitro

The germination of Moniliophthora perniciosa spores in rich media plates results in the growth of mycelia that presents a smooth, white aspect (Fig 8, left) with a continuous growth rate. After several months and plate passes, around ten, the mycelia begin to change its aspect becoming more rugged with a wave-like growth appearance (Fig 8, right) and the synthesis of a brown pigmentation. We consider the mycelia presented in the left panel of Fig 8 as young, and that in the right panel as old. This change appears to be irreversible, no regression to the young phenotype has been observed.

The yield of the DNA isolation depends greatly with the age of the mycelia used. Old mycelia render low amounts of material and present a relatively high degree of degradation. In contrast, young mycelia did not seem to show this problem. Another significant fact was that young mycelia had a higher mtDNA content with respect to genomic DNA than old mycelia, as estimated by the number of reads belonging to each class. For this reason we obtained many libraries encompassing different ages of the fungal culture. The consensus sequence obtained from the assembly of the reads belonging to young and old mycelia libraries did not show any difference.

Discussion

The mitochondrial genome of Moniliophthora perniciosa (109,103 bp) is the largest fungal mitochondrial genome
Fig 6 – Phylogenetic tree built from the aligned portions of concatenated protein sequences of cox1, cox2, and cox3. GenBank sequences used from: Allomyces macrogynus (NC_001715), Ashbya gossypii (NC_005789), Aspergillus niger (NC_007445), A. oryzae (NC_008282), A. tubingensis (NC_007597), Candida albicans (NC_002653), C. glabrata (NC_004691), C. metapsilosis (NC_006971), C. orthopsilosis (NC_006972), C. parapsilosis (NC_005789), C. zeminilina (NC_005972), C. neoformans (NC_004336), Debaryomyces hansenii (NC_010166), Epidermophyton floccosum (NC_007394) Gibberella zeae (NC_009493), Hanseniaspora uvarum (NC_007780), Harpochytrium sp. JEL105 (NC_004623), Harpochytrium sp. JEL94 (NC_004760), Hyaloraphidium curvatum (NC_003048), Hypocrea jecorina (NC_003388), Kluyveromyces lactis (NC_006077), K. thermotolerans (NC_006626), Lecanicillium muscarium (NC_004514), Metarhizium anisopliae (NC_008068), Moniliophthora perniciosa (NC_005927), Monoblepharella sp. JEL15 (NC_004624), Mortierella verticillata (NC_006838), Mycosphaerella graminicola (NC_010222), Paracoccidioides brasiliensis (NC_007935), Penicillium marneffei (NC_005256), Phaeosphaeria nodorum SN15 (NC_009746), Pichia canadensis (NC_001726), Pleurotus ostreatus (NC_009905), Podospora anserina (NC_001329), Rhizophyllum sp. 136 (NC_003053), Rhizopus oryzae (NC_006836), Saccharomyces castellii (NC_003920), S. cerevisiae (NC_001224), S. servazzii (NC_004918), Schizophyllum commune (NC_003049), Schizosaccharomyces japonicus (NC_004332), S. octosporus (NC_004312), S. pombe (NC_001326), Smittium culisetae (NC_006837), Spizellomyces punctatus (NC_003052), Tilletia indica (NC_009880), Ustilago maydis (NC_008368), Vanderwaltozyma polyssora (NC_009638), Verticillium dahliae (NC_003060) and Yarrowia lipolytica (NC_002659).
sequenced so far (http://www.ncbi.nlm.nih.gov.genomes/ORGANELLES/fu.html), the sixth among Basidiomycota, and the third of the class Agaricomycetes. To the best of our knowledge, the two members of the genus Moniliophthora, M. perniciosa and M. roreri, are the only ones from this taxonomic class with hemibiotrophic life-styles that also infect the aerial parts of plants. Therefore, the analysis of the fungal mitochondrial genomes sequenced so far (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/fu.html), the sixth among Basidiomycota, with potential to unveil specific information of this kind of organism.

This genome presents a circular topology, as have most of the fungal mitochondrial genomes sequenced so far. Only five, out of 50, have been described as linear. Also, it is A + T biased, like most other fungi. The tendency of A or T predominance becomes evident by the A/T frequency at the third codon position of the protein coding genes, 88.7 %, which is nearly a neutral position that could normally be occupied by any of the four bases without changing the amino acid identity. An explanation for this bias in the mitochondrial genome proposed by Xia (1996) suggest that, in order to maximize transcriptional efficiency, codon usage will be adapted to the most abundant nucleotide. Due to the high abundance of A + T, the presence of mitochondrial genes in the same orientation is an indication of the presence of one transcriptional unit. Transcription in mitochondria is universally accepted as policistronic (Schafer et al. 2005). Also, these ORFs and the known conserved genes are located in genomic regions not populated by tRNAs (Fig 1); so the protein coding regions are not superimposed by the tRNA coding regions, thus avoiding transcriptional problems. Conversely, the ORF coding for the tRNA without a defined anti-codon is located in a region populated by conserved genes (Fig 1, X at the innermost circle); this might indicate that the ORF is not real and does not code for anything. Transcription in mitochondria generates big transcripts that are processed with the tRNA regions indicating breaking points (tRNA punctuation model) in order to generate the final mRNAs (Ojala et al. 1981).

It is known that there are a large number of proteins in mitochondria, most of them encoded by nuclear genes (Dimmer & Rapaport 2008). Moreover, it is believed that most of these genes migrated from the mitochondrial genome.
to the nucleus. Therefore, it is possible that in fungi, which present a large variability in the size and structure of their mitochondrial genomes, some functional genes still remain in the mitochondrial genome contributing to the mitochondrial proteome.

With this information, together with the data presented here, we propose the existence of 17 potential new genes in the mitochondrial genome of *M. perniciosa* coding for proteins without known functions in this fungus. If this assumption is true, it is probable that the sequencing of new fungal mitochondrial genomes will lead to the identification of homologues.

**Phylogeny**

Concatenated sequences have been used successfully in phylogenetic analysis based on mitochondrial genes (Woo et al. 2003; Kouvelis et al. 2004). The phylogenetic analysis of the known fungal mitochondrial genomes is in accordance with recent studies using mitochondrial genes (Woo et al. 2003; Kouvelis et al. 2004), even considering the limited number of available complete genomes of *Basidiomycota*. Also, it concurs with the same type of analysis performed using genomic genes (James et al. 2006).

The phylogenetic tree built with the polymerase genes present in the plasmid region show a completely different picture. The species segregates into two groups according to the type of plasmid, kalilo-like and non-kalilo-like (Fig 7, right and left, respectively). Both groups have members of the *Ascomycota* and *Basidiomycota* not following the phylogenetic analysis using the cox mitochondrial genes (Woo et al. 2003; Kouvelis et al. 2004) or genomic genes (James et al. 2006). The most significant finding is the segregation of the plasmids in two groups. This segregation suggests an ancient duplication of these plasmids, before the separation of *Ascomycota* and *Basidiomycota*, following a separate evolutionary path independently of the host.

**Synteny**

Fungal mitochondrial genomes do not show an overall high degree of synteny (Kouvelis et al. 2004), in contrast to the metazoan mitochondrial genomes that show the same gene content and order (Boore 1999). Significant synteny has been described for species belonging to the classes *Sordariomycetes* (Kouvelis et al. 2004) and *Eurotiomycetes* (Cardoso et al. 2007). These data indicate that, at least at the level of class, some fungal mitochondrial genomes from *Ascomycota* show significant synteny. In the case of *Basidiomycota*, the number of sequenced mitochondrial genomes is significantly lower, but out of six genomes three of them belong to the class *Agaricomycetes*. The analysis of the content and order of genes did not show an overall synteny in this case, in contrast to what happens in the case of the *Sordariomycetes* and *Eurotiomycetes*, which have clear synteny. Although we have very limited data, they seem to indicate that some groups of *Basidiomycota* are more diverse than some groups of *Ascomycota*.

Conversely, we have recently sequenced the complete mitochondrial genome of *Moniliophthora rorera*, a species very close to *M. perniciosa*. A preliminary analysis showed that, in this case, all known conserved genes present the same order in the two genomes (result not shown) indicating a high degree of synteny. Taken together, our data indicate that synteny is not a general feature of the *Basidiomycota* mitochondrial genome and that possibly gene order is mixed during speciation.

**Mitochondrial plasmid**

Many filamentous fungi have mitochondrial plasmids (Griffiths 1995). These plasmids do not appear to have any phenotypic effect on their hosts with a few exceptions (Bertrand 2000; Griffiths 1992). These include the mitochondrial plasmids pKALILO of *Neurospora intermedia* (Bertrand et al. 1985, 1986), pMARANHAR of *N. crassa* (Court et al. 1991), and pAL2-1 of *Podospora anserina* (Maas et al. 2004). These plasmids have a typical invertron structure (with long inverted repeats) and affect the life span of their hosts. *Neurospora* strains normally do not senesce, but those carrying one of the mitochondrial plasmids pKALILO or pMARANHAR die in two to three weeks. This is associated with mitochondrial instability and insertions of the plasmid into the mitochondrial genome (Bertrand et al. 1985, 1986; Myers et al. 1989; Rieck et al. 1982), and for these mitochondrial plasmids have been named senescence plasmids (Griffiths 1992). In the naturally senescent species *P. anserina*, insertion of pAL2-1 in strain AL2 is correlated with an increased life-span, and therefore, this mitochondrial plasmid has been named life-span-prolonging plasmid (Hermanns et al. 1994). However, this life-span extension requires more than the presence of pAL2-1, as the majority of natural isolates containing pAL2-1 homologous plasmids do not show this effect (van der Gaag et al. 1998). This process seems to be a combination of the insertion of the plasmid and a complex set of mitochondrial DNA rearrangements, including two insertions of the plasmid and a 3.6 kb deletion of the region between the cyt*b and atp6 genes (Osiewacz et al. 1989; Hermanns & Osiewacz 1992). This process may have stabilized the plasmid sequence as an integral part of a mitochondrial genome, which thereby may have acquired new functions.

We detected possible sequence rearrangements in the neighbourhood of the plasmid region of the *Moniliophthora perniciosa* mitochondrial genome. Typically, the length of inverted repeats found in linear plasmids is around 1000 bp (Griffiths 1998). The plasmid region in *M. perniciosa* shows four sets of inverted repeats, two smaller (130 bp) and two larger (347 bp), and three conserved hypothetical ORFs located between the smaller and the larger repeats (Fig 4A). It seems that these ORFs were inserted inside the inverted repeats disrupting them, creating the new four smaller inverted repeats. Also, four non-conserved, hypothetical ORFs are located just after the last of the larger inverted repeats in the same region of the genome that is defined by its lower GC content and a change in direction of the GC skew (Fig 3). It is interesting to note that two of the conserved hypothetical ORFs, *hylF1* and *hylF3*, present a significant degree of similarity and they are located at the inner part of the extremities of the region delimited by the larger inverted repeats, their coding region starts just inside the inverted repeats, and they present opposite orientation. Also, the non-conserved hypothetical ORFs
hyp58 shows significant similarity with hypP1 and hypP3. There are two more non-conserved hypothetical ORFs, hyp57 and hyp28, with opposite orientation and significant similarity. The existence of ORFs between the inverted repeats is an indication of an integration–recombination event that took place in this region. We believe that this event disrupted the typical inverted repeats present in linear plasmids, and as a consequence, the plasmid became stably integrated and remained part of the genome. Also, the non-conserved, hypothetical ORFs with significant similarity and opposite orientation indicate the occurrence of such integration–recombination events at this region of the genome.

Some symptoms that characterize fungal ageing are a decrease of growth rate and fertility, a reduction in the formation of aerial hyphae, and an increase in pigmentation. Except for the fertility, which has not been studied, all of the other symptoms have been observed in M. perniciosa growing in the laboratory (Fig 8). We are currently studying this process in order to see whether it can be characterized as senescence. Ageing in Neurospora is linked to the process of integration–excision of the linear plasmids (Osiewacz et al. 1989; Hermanns & Osiewacz 1992). Our sequencing data showed that the plasmid in M. perniciosa is stably integrated, thus, the origin of the ageing process observed in this fungal species seems to be different from that provoked by the linear plasmids in Neurospora.

The plasmid exists in all the isolates of M. perniciosa analysed indicating that it is an integral part of its mitochondrial genome. The GC content and GC skew indicate that the integration of the plasmid is a recent event and that the sequence is not adapted to the rest of the mitochondrial genome, thus maintaining some of its original features. It would be interesting to know the behaviour of the species before the integration of this element in its mitochondrial genome.

In our survey of many isolates of M. perniciosa from diverse geographical origins, we consistently found the plasmid integrated in the mitochondrial genome and concluded that this is a feature of the species, mitochondrial genome. Remarkably, a preliminary analysis of the mitochondrial genome of the close species M. roreri showed a high degree of similarity between both genomes (a similarity of 57% for the comparison of M. roreri with M. perniciosa) but, curiously, the plasmid region was not present (not shown).

Although speculative, we believe that the insertion could confer some evolutionary advantage to the species. If this was not the case, the insertion would have suffered a negative selection or, as happens in P. anserina, we would have two types of isolates: one without the integrated plasmid and one with it in the mitochondrial genome.

In summary, we described in this work the largest basidiomycete mitochondrial genome sequenced so far. This genome has some remarkable features, such as the presence of putative new mitochondrial genes and the stable integration of a complete linear plasmid sequence. We are currently investigating how these elements can contribute to the physiology of this fungus.

Acknowledgements

This research was supported by the Brazilian agencies CNPq (research fellowship to N. C.), Capes, CNPq Regional Genoma Program, SEAGRIm and FAPESP (No. 02/09280-1).

REFERENCES


The Moniliophthora perniciosa mitochondrial genome


Myers CJ, Griffths AJF, Bertrand H, 1989. Linear kalilo DNA is a Neurospora mitochondrial plasmid that integrates into the mitochondrial DNA. Molecular and General Genetics 220: 113–120.


